

(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2018-45537
(P2018-45537A)

(43) 公開日 平成30年3月22日(2018.3.22)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 17/30 (2006.01)	G06F 17/30 220Z	
	G06F 17/30 170A	
	G06F 17/30 320D	

審査請求 未請求 請求項の数 7 O L (全 27 頁)

(21) 出願番号	特願2016-181003 (P2016-181003)	(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9月15日(2016.9.15)	(74) 代理人	110002147 特許業務法人酒井国際特許事務所
		(72) 発明者	高瀬 信子 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	嶺野 和夫 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	伊藤 尚洋 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

(54) 【発明の名称】 検索プログラム、検索装置および検索方法

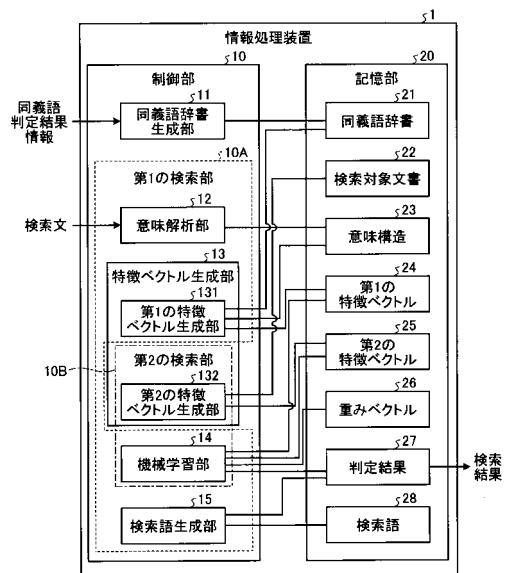
(57) 【要約】

【課題】 検索対象文書から特定の文章を検索する場合に、検索結果に含まれるノイズを低減する。

【解決手段】 情報処理装置1は、検索対象文書から特定の文章を検索する場合に、文章を受け付け、受け付けた文章を意味解析して、受け付けた文章に含まれる単語の意味を示す意味構造23を生成し、単語と当該単語の意味を示す意味構造23とを対応付けて記憶する同義語辞書21を参照して、生成した意味構造23に対応付けられた単語を特定し、特定した単語が検索対象文書22に含まれるか否かの判定を行い、判定結果に応じた情報を出力する。

【選択図】 図1

実施例に係る情報処理装置の構成の一例を示す機能ブロック図



【特許請求の範囲】**【請求項 1】**

文章データから特定の文章を検索する検索プログラムにおいて、
文章を受け付け、
受け付けた前記文章を意味解析して、受け付けた前記文章に含まれる単語の意味を示す情報を生成し、
単語と当該単語の意味を示す情報とを対応付けて記憶する記憶部を参照して、生成した前記情報に対応付けられた単語を特定し、
特定した前記単語が前記文章データに含まれるか否かの判定を行い、
判定結果に応じた情報を出力する
処理をコンピュータに実行させることを特徴とする検索プログラム。

10

【請求項 2】

前記単語が前記文章データに含まれる場合には、前記文章データに含まれる複数の単語それぞれを該複数の単語それぞれの意味を示す情報に対応付けて記憶する第 2 の記憶部を参照して、生成した前記情報に対応付けられた単語を特定し、
特定した前記単語が前記文章データの中のいずれの文章に含まれるか否かの判定を行い、
判定結果に応じた情報を出力する
ことを特徴とする請求項 1 に記載の検索プログラム。

20

【請求項 3】

前記第 2 の記憶部を参照して特定する処理は、前記第 2 の記憶部を参照して、前記記憶部を参照して特定した単語と一致する単語に対応付けられた意味を示す情報を特定し、特定した前記情報と、該生成した前記情報とを用いて、該生成した前記情報に対応付けられた単語を特定する
ことを特徴とする請求項 2 に記載の検索プログラム。

【請求項 4】

前記意味を示す情報は、該当する文章に含まれる単語と、他の単語との関係を示す情報を含む
ことを特徴とする請求項 1 から請求項 3 のいずれか 1 つに記載の検索プログラム。

30

【請求項 5】

該生成する処理は、受け付けた前記文章の前記意味を示す情報から、前記文章に含まれる前記単語と、前記単語と直接関係する他の単語との関係を示す情報を抽出し、抽出した情報を前記単語の意味を示す情報として生成する
ことを特徴とする請求項 4 に記載の検索プログラム。

【請求項 6】

文章データから特定の文章を検索する検索装置において、
文章を受け付ける受付部と、
前記受付部によって受け付けられた前記文章を意味解析して、受け付けた前記文章に含まれる単語の意味を示す情報を生成する生成部と、
単語と当該単語の意味を示す情報とを対応付けて記憶する記憶部を参照して、生成した前記情報に対応付けられた単語を特定する特定部と、
前記特定部によって特定された前記単語が前記文章データに含まれるか否かの判定を行う判定部と、
前記判定部によって判定された結果である判定結果に応じた情報を出力する出力部と、
を有することを特徴とする検索装置。

40

【請求項 7】

文章データから特定の文章を検索する検索方法において、
文章を受け付け、
受け付けた前記文章を意味解析して、受け付けた前記文章に含まれる単語の意味を示す情報を生成し、

50

単語と当該単語の意味を示す情報とを対応付けて記憶する記憶部を参照して、生成した前記情報に対応付けられた単語を特定し、

特定した前記単語が前記文章データに含まれるか否かの判定を行い、判定結果に応じた情報を出力する

各処理をコンピュータが実行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、検索プログラムなどに関する。

【背景技術】

【0002】

近年、ネットワークの発達により増大している情報量の情報の中から目的の情報が高精度に検索されることが求められている。

【0003】

情報量が少ない場合には、ノイズを含んだ（意図しない）検索結果が出力されても、利用者が検索結果を吟味して目的の文章を探し出すことができる。ところが、情報量が多くなると、検索結果が多くなり、利用者が検索結果を吟味することは困難である。このため、ノイズを減らして検索結果が出力されることが求められる。

【0004】

検索手法の一例を、図18を参照して説明する。図18は、検索手法の一例である。図18に示すように、情報を検索する検索装置は、検索文として文章が入力された場合に、入力された文章を形態素解析し、単語単位に分割する（S100）。そして、検索装置は、分割された単語を、予め同義語を記憶した同義語辞書と照合し（S110）、一致した単語を含む同義語を検索語として生成する（S120）。かかる同義語辞書は、単語単位の表記で意味が同じと思われる単語同士を同義語として格納する。そして、検索装置は、検索対象文書から検索語が含まれる文章を検索結果として検索する（S130、S140）。

【0005】

例えば、図18では、検索文として「画像をクリアに表示するには」という文章が入力されたとする。すると、検索装置は、入力された文章を形態素解析し、単語単位に分割する。この結果、分割された単語は、「画像」、「クリア」、「表示」を示す。そして、検索装置は、分割された単語を、同義語辞書と照合し、一致した単語を含む同義語を検索語として生成する。同義語辞書に「クリア、消去」、「表示、表現」が同義語として含まれるとする。すると、検索語として、「画像」と、「クリア」または「消去」と、「表示」または「表現」が生成される。そして、検索装置は、検索対象文書から検索語が含まれる文章を検索結果として検索する。この結果、検索装置は、検索結果として、「画像データをクリアするには・・・」「過去の画像の表示を消去して・・・」「正規表現をいったんクリア・・・」の文章を検索する。

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開2015-60243号公報

【発明の概要】

【発明が解決しようとする課題】

【0007】

しかしながら、従来の検索手法では、検索対象文書から特定の文章を検索する場合に、検索結果に含まれるノイズを低減することができないという問題がある。例えば、図18で示す従来の検索手法では、検索語として、検索文の「クリア」とは違う意味の「消去」や、検索文の「表示」とは違う意味になり得る「表現」が含まれる。このため、検索装置は、検索対象文書から特定の文章を検索するときに、意味が違う検索語が含まれる文章を

10

20

30

40

50

検索結果として検索してしまう。すなわち、検索文の「クリア」とは違う意味の「クリア」や「消去」、検索文の「表現」とは違う意味の「表現」を含んだ文章を検索結果として検索してしまう。この結果、検索装置は、検索結果に含まれるノイズを低減することができない。

【0008】

本発明は、1つの側面では、検索対象文書から特定の文章を検索する場合に、検索結果に含まれるノイズを低減することを目的とする。

【課題を解決するための手段】

【0009】

1つの態様では、検索プログラムは、文章データから特定の文章を検索する検索プログラムにおいて、文章を受け付け、受け付けた前記文章を意味解析して、受け付けた前記文章に含まれる単語の意味を示す情報を生成し、単語と当該単語の意味を示す情報とを対応付けて記憶する記憶部を参照して、生成した前記情報に対応付けられた単語を特定し、特定した前記単語が前記文章データに含まれるか否かの判定を行い、判定結果に応じた情報を出力する、処理をコンピュータに実行させる。

10

【発明の効果】

【0010】

1実施態様によれば、検索結果に含まれるノイズを低減することができる。

【図面の簡単な説明】

【0011】

20

【図1】図1は、実施例に係る情報処理装置の構成の一例を示す機能ブロック図である。

【図2】図2は、実施例に係る同義語辞書の一例を示す図である。

【図3A】図3Aは、意味構造の一例を示す図(1)である。

【図3B】図3Bは、意味構造の一例を示す図(2)である。

【図3C】図3Cは、用語の意味構造の一例を示す図である。

【図4】図4は、実施例に係る特徴ベクトルの構成の一例を示す図である。

【図5】図5は、重みベクトルの導出の一例を示す図である。

【図6A】図6Aは、実施例に係る第1の検索処理の一例を示す図(1)である。

【図6B】図6Bは、実施例に係る第1の検索処理の一例を示す図(2)である。

【図6C】図6Cは、実施例に係る第1の検索処理の一例を示す図(3)である。

30

【図6D】図6Dは、実施例に係る第1の検索処理の一例を示す図(4)である。

【図6E】図6Eは、実施例に係る第1の検索処理の一例を示す図(5)である。

【図6F】図6Fは、実施例に係る第1の検索処理の一例を示す図(6)である。

【図7】図7は、実施例に係る第1の検索処理の流れを示す図である。

【図8A】図8Aは、実施例に係る第2の検索処理の一例を示す図(1)である。

【図8B】図8Bは、実施例に係る第2の検索処理の一例を示す図(2)である。

【図9】図9は、実施例に係る第2の検索処理の流れを示す図である。

【図10】図10は、実施例に係る同義語辞書生成処理のフローチャートの一例を示す図である。

【図11】図11は、同義語判定結果情報のデータ構造の一例を示す図である。

40

【図12】図12は、実施例に係る検索処理のフローチャートの一例を示す図である。

【図13】図13は、実施例に係る第1の特徴ベクトル生成処理のフローチャートの一例を示す図である。

【図14】図14は、実施例に係る第2の特徴ベクトル生成処理のフローチャートの一例を示す図である。

【図15】図15は、実施例に係る機械学習処理のフローチャートの一例を示す図である。

【図16】図16は、実施例に係る検索語生成処理のフローチャートの一例を示す図である。

【図17】図17は、検索プログラムを実行するコンピュータの一例を示す図である。

50

【図 1 8】図 1 8 は、検索手法の一例を示す図である。

【発明を実施するための形態】

【0012】

以下に、本願の開示する検索プログラム、検索装置および検索方法の実施例を図面に基
づいて詳細に説明する。なお、実施例では、検索装置を情報処理装置として説明する。本
発明は、実施例により限定されるものではない。

【実施例】

【0013】

[実施例に係る情報処理装置の構成]

図 1 は、実施例に係る情報処理装置の構成の一例を示す機能ブロック図である。図 1 に
示す情報処理装置 1 は、機械翻訳などで利用される自然言語の意味解析処理の結果（「意
味構造」という）を用いて、2 段階の機械学習で評価を行い、複数の検索対象文を含む検
索対象文書から検索文と意味が同じ用語を含む検索対象文を抽出する。1 段階目の機械学
習では、情報処理装置 1 は、検索文の用語に対する意味構造と、同義語辞書の用語に対す
る意味構造とを組み合わせる評価し、同義語辞書から検索文に含まれる用語と意味が同じ
同義語を検索語として抽出する。2 段階目の機械学習では、情報処理装置 1 は、検索語に
より検索対象文書の用語を検索し、ヒットした検索対象文の用語に対する意味構造と、検
索文の用語に対する意味構造とを組み合わせる評価する。そして、情報処理装置 1 は、多
義語を含む検索対象文を排除し、検索文と意味が同じ用語を含む検索対象文を複数の検
索対象文書から抽出する。「同義語」とは、異なる表記であるが同じ意味を表す表記の用語
同士のことをいう。「多義語」とは、同じ表記であるが別の意味を表す表記の用語同士の
ことをいう。

10

20

【0014】

なお、実施例で用いられる「文章」とは、意味のある文の最小単位であり、必ずしも主
語および述語を含まなくても良いが、日本語では一般的に句点（。）で区切られ、英語で
はピリオド（.）で区切られる。「文章」とは、「文」と同義である。「文書」とは、複
数の文章として構成されるテキストのことをいう。また、実施例で用いられる「意味構造
」とは、文章の意味解析処理の結果として生成され、文章の意味を表す、概念（意味）を
示すノードと、概念の関連を示す有向アークとで表わしたグラフ構造のことをいう。また
、実施例で用いられる「意味属性」とは、「意味構造」のノードについて、文法的、意味
的な性質を示す記号（属性）のことをいう。また、実施例で用いられる用語とは、単語の
一例であるとする。

30

【0015】

情報処理装置 1 は、制御部 10 と記憶部 20 とを有する。

【0016】

制御部 10 は、CPU (Central Processing Unit) などの電子回路に対応する。そ
して、制御部 10 は、各種の処理手順を規定したプログラムや制御データを格納するた
めの内部メモリを有し、これらによって種々の処理を実行する。制御部 10 は、同義語辞書
生成部 11 と、意味解析部 12 と、特徴ベクトル生成部 13 と、機械学習部 14 と、検索
語生成部 15 とを有する。特徴ベクトル生成部 13 は、第 1 の特徴ベクトル生成部 13 1
と、第 2 の特徴ベクトル生成部 13 2 とを有する。なお、意味解析部 12、第 1 の特徴ベ
クトル生成部 13 1、機械学習部 14 および検索語生成部 15 は、第 1 の検索部 10 A に
含まれる。第 2 の特徴ベクトル生成部 13 2 および機械学習部 14 は、第 2 の検索部 10
B に含まれる。なお、第 1 の検索部 10 A は、1 段階目の機械学習に対応し、第 2 の検
索部 10 B は、2 段階目の機械学習に対応する。

40

【0017】

記憶部 20 は、例えば、RAM、フラッシュメモリ (Flash Memory) などの半導体メ
モリ素子、または、ハードディスク、光ディスクなどの記憶装置である。記憶部 20 は、
同義語辞書 21、検索対象文書 22、意味構造 23、第 1 の特徴ベクトル 24、第 2 の特
徴ベクトル 25、重みベクトル 26、判定結果 27 および検索語 28 を有する。なお、意

50

味構造 2 3 は、意味を示す情報の一例である。

【 0 0 1 8 】

検索対象文書 2 2 は、複数の検索対象文を含む文書である。検索対象文書 2 2 は、文書に含まれる検索対象文ごとに、意味構造 2 3 を対応付けて記憶する。

【 0 0 1 9 】

同義語辞書 2 1 は、意味が同じと判定される用語同士を同義語としてグループ化した辞書である。同義語辞書 2 1 は、同義語として判定されたそれぞれの用語を、それぞれの用語を意味で判定できる範囲の意味構造 2 3 と対応付けて記憶する。

【 0 0 2 0 】

ここで、実施例に係る同義語辞書 2 1 の一例を、図 2 を参照して説明する。図 2 は、実施例に係る同義語辞書の一例を示す図である。図 2 に示すように、同義語辞書 2 1 は、フラグ 2 1 a と、用語の表記 2 1 b と、意味構造 2 1 c とを対応付けて記憶する。

10

【 0 0 2 1 】

フラグ 2 1 a は、1 つの同義語に含まれる用語の区別を示すフラグである。1 つの同義語に含まれる一方の用語を示すフラグを用語 1 とし、他方の用語を示すフラグを用語 2 とする。なお、1 つの同義語に含まれる用語は、2 つに限定されず、3 つ以上であっても良い。例えば、3 つの場合には、フラグを、用語 1、用語 2 および用語 3 とすれば良い。

【 0 0 2 2 】

用語の表記 2 1 b は、同義語に含まれる用語の表記を表す。意味構造 2 1 c は、用語の表記 2 1 b で示される用語の意味を判定できる範囲の意味構造 2 3 である。言い換えれば、意味構造 2 1 c は、用語の表記 2 1 b で示される用語を含む文章の意味構造 2 3 から当該用語の意味を判定できる範囲を切り出したものである。なお、以降、用語を意味で判定できる範囲の意味構造 2 3 のことを「用語の意味構造」というものとする。

20

【 0 0 2 3 】

一例として、フラグ 2 1 a が「用語 1」である場合に、用語の表記 2 1 b として「一意」、意味構造 2 1 c として「z 1」を記憶する。フラグ 2 1 a が用語 2 である場合に、用語の表記 2 1 b として「ユニーク」、意味構造 2 1 c として「z 2」を記憶する。つまり、「一意」と「ユニーク」とが同義であることを示す。フラグ 2 1 a が用語 1 である場合に、用語の表記 2 1 b として「クリア」、意味構造 2 1 c として「x 1」を記憶する。フラグ 2 1 a が「用語 2」である場合に、用語の表記 2 1 b として「クリア」、意味構造 2 1 c として「x 2」を記憶する。つまり、「クリア」と「クリア」とが同義であることを示す。

30

【 0 0 2 4 】

ここで、意味構造 2 3 の一例を、図 3 A、図 3 B および図 3 C を参照して説明する。図 3 A および図 3 B は、意味構造の一例を示す図である。図 3 A は、意味構造 2 3 の内部表現を示す図であり、図 3 B は、意味構造 2 3 を図式化したものである。また、図 3 C は、用語の意味構造の一例を示す図である。なお、図 3 A および図 3 B で示す意味構造 2 3 は、原文「運用環境のカスタマイズが必要です。」の意味解析処理の結果である。

【 0 0 2 5 】

図 3 A に示すように、意味構造 2 3 の内部表現は、用語一覧とグラフ構造を含む。

40

【 0 0 2 6 】

用語一覧には、用語の一覧が表わされる。1 つの用語は、「表記」「品詞」「意味属性」「概念記号」で表わされる。「意味属性」は、対応する用語の文法的、意味的な性質を示す記号（属性）であり、例えば、シソーラス属性や、可算規則、変化形規則などがある。シソーラス属性とは、用語の意味的な階層関係を記述した属性のことをいう。「概念記号」とは、意味的な観点（概念レベル）において単語が指し示す概念を識別する記号のことをいう。用語の一例として、表記が「カスタマイズ」である場合に、品詞としてサ変名詞「SN」、意味属性として「S 1, S 2, …」、概念記号として「CUSTOMIZE」と記載されている。

【 0 0 2 7 】

50

グラフ構造は、(Fromノード) - - <アーク> - - > (Toノード)で表現され、(Fromノード)の用語から(Toノード)の用語に向かって、<アーク>で接続されていることを示す。各ノードには、用語の概念記号が付与されている。一例として、(CUSTOMIZE) - - <OBJ> - - > (EJR01)について、ノードが示す概念記号(CUSTOMIZE)の用語「カスタマイズ」とアーク<OBJ>で接続するノードが概念記号(EJR01)の用語「運用環境」であることを示す。各ノードは、用語一覧に示されている「表記」「品詞」「意味属性」「概念記号」を持っている。

【0028】

図3Bに示すように、例えば、「カスタマイズ」という用語は、「CUSTOMIZE」という概念記号で表わされている。括弧内の記号が概念記号である。そして、「CUSTOMIZE」という概念記号のノードは、「EJR01」という概念記号で表わされる「運用環境」という用語と、「OBJ」という記号を持つアークで接続されている。さらに、「CUSTOMIZE」という概念記号のノードには、他方のノードが存在しないアークが接続されている。このアークは、「J.GA」という記号を示す。また、「CUSTOMIZE」という概念記号のノードは、「NEED」という概念記号で表わされる「必要」という用語と、「OBJA」という記号を持つアークで接続されている。つまり、意味構造23は、用語と用語の意味的な関係を表す有向グラフで表される。

10

【0029】

図3Cに示す意味構造23は、用語1「カスタマイズ」の意味構造23である。かかる「カスタマイズ」の意味構造23は、「カスタマイズ」の「概念記号」「品詞」「意味属性」と、「カスタマイズ」の前後1ノード(NULLのノードを含む)とノード間を結ぶアークとを含む。ここでは、用語1「カスタマイズ」の概念記号は、「CUSTOMIZE」であり、品詞は、「SN」であり、意味属性は、「S1, S2, ...」である。「カスタマイズ」の前1ノードは、概念記号を「NEED」として表される用語「必要」のノードであり、「カスタマイズ」の後1ノードは、概念記号を「EJR01」として表される用語「運用環境」のノードである。そして、「カスタマイズ」と「必要」との間を結ぶアークは、「OBJA」で表わされ、「カスタマイズ」と「運用環境」との間を結ぶアークは、「OBJ」で表わされる。また、「カスタマイズ」とNULLのノードとの間を結ぶアークは、「J.GA」で表わされる。すなわち、用語1「カスタマイズ」の意味構造23は、「運用環境のカスタマイズが必要です。」という原文の意味構造23から用語1「カスタマイズ」の意味を判定できる範囲を切り出したものである。

20

30

【0030】

図1に戻って、同義語辞書生成部11は、意味構造23を含む同義語辞書21を生成し、記憶部20に格納する。例えば、同義語辞書生成部11は、同義語判定結果情報に設定された同義語ごとに、同義語の各用語(用語1、用語2)について、意味構造23を生成する。ここでいう同義語判定結果情報とは、特定分野の複数の文章に含まれる複数の用語を対象にして、同義語か否かが判定された各用語(用語1、用語2)と、各用語が出現する文章の行番号と、同義語か否かを示すフラグとを予め設定したものである。なお、同義語判定結果情報は、各用語が出現する文章の行番号の代わりに、各用語が出現する文章の意味構造23の行番号としても良い。

40

【0031】

一例として、同義語辞書生成部11は、同義語判定結果情報の同義語と判定された各用語(用語1、用語2)が出現する行番号に対応するそれぞれの文章について、自然言語の意味解析処理により、それぞれの意味構造23を生成する。そして、同義語辞書生成部11は、用語1が出現する文章の意味構造23から用語1の周辺を分離した、用語1の意味構造23を生成する。すなわち、同義語辞書生成部11は、用語1が出現する文章の意味構造23から用語1の意味を判定できる範囲を切り出し、用語1の意味構造23を生成する。同様に、同義語辞書生成部11は、用語2が出現する文章の意味構造23から用語2の周辺を分離した、用語2の意味構造23を生成する。すなわち、同義語辞書生成部11は、用語2が出現する文章の意味構造23から用語2の意味を判定できる範囲を切り出し

50

、用語 2 の意味構造 2 3 を生成する。そして、同義語辞書生成部 1 1 は、用語 1 および用語 2 について、各用語のフラグ、各用語の表記、各用語の意味構造 2 3 を同義語辞書 2 1 に書き出す。

【0032】

また、同義語辞書生成部 1 1 は、同義語辞書 2 1 の中で、用語 1 の表記および意味構造 2 3 と用語 2 の表記および意味構造 2 3 とが同じものを削除する。同義語辞書生成部 1 1 は、同義語辞書 2 1 の中で、用語 1 と用語 2 のペアが、他の用語 1 と用語 2 のペアと同じ場合、どちらか一方のペアを削除する。

【0033】

意味解析部 1 2 は、検索対象の検索文を意味解析する。なお、意味解析部 1 2 は、第 1 の検索部 1 0 A に含まれる。例えば、意味解析部 1 2 は、検索対象の検索文について、形態素解析および意味解析を行い、意味構造 2 3 を生成する。すなわち、意味解析部 1 2 は、検索対象の検索文について、自然言語の意味解析処理により意味構造 2 3 を生成する。

10

【0034】

なお、同義語辞書生成部 1 1 によって行われる意味解析処理および意味解析部 1 2 によって行われる意味解析処理は、既存の機械翻訳技術を使って実現できる。例えば、かかる意味解析処理は、例えば、特開平 6 - 6 8 1 6 0 号公報、特開昭 6 3 - 1 3 6 2 6 0 号公報や特開平 4 - 3 7 2 0 6 1 号公報に開示されている機械翻訳技術を使って行えば良い。また、意味構造 2 3 は、例えば、特開 2 0 1 2 - 7 3 9 5 1 号公報に開示されている。

【0035】

第 1 の特徴ベクトル生成部 1 3 1 は、検索文の用語と、同義語辞書 2 1 の用語とを組み合わせ、機械学習で用いられる第 1 の特徴ベクトル 2 4 を生成する。なお、第 1 の特徴ベクトル生成部 1 3 1 は、第 1 の検索部 1 0 A に含まれる。

20

【0036】

例えば、第 1 の特徴ベクトル生成部 1 3 1 は、検索文の用語の意味構造 2 3 を生成する。用語の意味構造 2 3 の生成方法は、同義語辞書生成部 1 1 で用いた方法と同様である。すなわち、第 1 の特徴ベクトル生成部 1 3 1 は、意味解析部 1 2 によって生成された検索文の意味構造 2 3 から検索文の用語の周辺を分離した、用語の意味構造 2 3 を生成する。そして、第 1 の特徴ベクトル生成部 1 3 1 は、生成した検索文の用語の意味構造 2 3 から、特徴ベクトルを生成する。

30

【0037】

ここでいう特徴ベクトルとは、用語の意味構造 2 3 の情報を素性とする特徴情報のことである。例えば、特徴ベクトルは、用語を示す情報（品詞、意味属性および概念記号）と、当該用語と直接アークで接続する用語を示す情報（品詞、意味属性および概念記号）とを素性とする特徴情報である。なお、特徴ベクトルの詳細は、後述する。

【0038】

また、第 1 の特徴ベクトル生成部 1 3 1 は、同義語辞書 2 1 から、検索文の用語と表記が一致した用語と同じグループ内の当該用語と異なる用語を抽出する。つまり、第 1 の特徴ベクトル生成部 1 3 1 は、同義語辞書 2 1 から、検索文の用語と意味が同じである可能性がある用語を抽出する。そして、第 1 の特徴ベクトル生成部 1 3 1 は、同義語辞書 2 1 において、抽出した用語に対応付けられた意味構造 2 1 c を取得する。そして、第 1 の特徴ベクトル生成部 1 3 1 は、取得した意味構造 2 1 c から、特徴ベクトルを生成する。なお、特徴ベクトルの構成は、後述する。

40

【0039】

ここで、特徴ベクトルの構成の一例を、図 4 を参照して説明する。図 4 は、実施例に係る特徴ベクトルの構成の一例を示す図である。なお、図 4 で示す特徴ベクトルは、対象の用語に対応する対象ノードの情報 $i 0$ と、対象の用語と直接アークで接続する用語に対応する関連ノードの情報 $i 1 0$ とを素性とする特徴情報の一例である。図 4 に示すように、特徴ベクトルには、対象の用語に対応する対象ノードに対して、各種品詞の素性、各種意味属性の素性および各種概念記号の素性が含まれる。各種品詞の素性は、例えば、30 種

50

存在する。各種意味属性の素性は、例えば、10000種存在する。そして、各種品詞の素性の中で対象の用語の素性と合致する素性の値として1が設定される。各種意味属性の素性の中で判定すべき用語の素性と合致する素性の値として1が設定される。各種概念記号の素性の中で判定すべき用語の素性と合致する素性の値として1が設定される。加えて、特徴ベクトルには、対象の用語と直接アークで接続する用語に対応する対象ノードに対して、各種アークの数分の、方向の素性、各種品詞の素性、各種意味属性の素性および各種概念記号の素性が含まれる。各種アークは、例えば、700種存在する。方向の素性は、1種存在する。各種品詞の素性は、例えば、30種存在する。各種意味属性の素性は、例えば、10000種存在する。そして、対象ノードが直接アークで接続する関連ノードの情報、当該アークに対応する領域に設定される。各種品詞の素性の中で、関連ノードに対応する用語の素性と合致する素性の値として1が設定される。各種意味属性の素性の中で、関連ノードに対応する用語の素性と合致する素性の値として1が設定される。各種概念記号の素性の中で、関連ノードに対応する用語の素性と合致する素性の値として1が設定される。なお、図4で示す特徴ベクトルは、品詞について、先頭の素性をADJとし、意味属性について、素性の順番をS1、S2、S3としたが、これに限定されず、予め定められた素性の順番であれば良い。

10

【0040】

図1に戻って、第1の特徴ベクトル生成部131は、生成された2種の特徴ベクトルを比較し、比較結果の特徴ベクトルを生成する。一例として、第1の特徴ベクトル生成部131は、生成された2種の特徴ベクトルについて、先頭の素性から順次素性の値を比較し、一致する素性の値を1とし、一致しない値を0とする比較結果の特徴ベクトルを生成する。

20

【0041】

そして、第1の特徴ベクトル生成部131は、生成された3種の特徴ベクトル、すなわち、2種の意味構造23からそれぞれ生成された2種の特徴ベクトルと比較結果の特徴ベクトルとを連結して機械学習用の第1の特徴ベクトル24を生成する。このような第1の特徴ベクトル24を用いて、同じ意味を持つ可能性が高い用語同士が同じ意味か否かが機械学習器により評価可能になる。

【0042】

機械学習部14は、第1の特徴ベクトル24を教師あり学習を行う機械学習器で評価し、検索文の用語と、同義語辞書21から抽出された用語とが同義語であるか否かを判定する。なお、ここでいう機械学習器には、例えば、サポートベクトルマシン(SVM)が挙げられる。以降では、機械学習器としてSVMを採用した場合について説明する。例えば、機械学習部14は、第1の検索部10Aの場合には、第1の特徴ベクトル24と予め教師例に基づいて学習された重みベクトル26との内積を計算し、総合評価値を導出する。そして、機械学習部14は、導出した総合評価値を予め設定された閾値で判定することにより、同義語であるか否かの判定結果27を決定する。

30

【0043】

また、機械学習部14は、後述する第2の特徴ベクトル25を教師あり学習を行う機械学習器で評価し、検索文の用語と、検索対象文書22から抽出された用語とが同義語であるか否かを判定する。例えば、機械学習部14は、第2の検索部10Bの場合には、第2の特徴ベクトル25と、予め教師例に基づいて学習された重みベクトル26との内積を計算し、総合評価値を導出する。そして、機械学習部14は、導出した総合評価値を予め設定された閾値で判定することにより、同義語であるか否かの判定結果27に基づいて検索結果を出力する。

40

【0044】

検索語生成部15は、第1の検索部10Aにおいて、検索文の用語と、検索文の用語と同義語であると判定された用語とを検索語28として生成する。例えば、検索語生成部15は、検索文の用語を検索語28とする。検索語生成部15は、機械学習部14によって決定された判定結果27の中から同義語と判定された判定結果27を取得する。検索語生

50

成部 15 は、取得した判定結果 27 を有する、同義語辞書 21 から抽出された用語を検索語 28 とする。そして、検索語生成部 15 は、検索語 28 とされた用語の中で、用語の表記が重複している場合には、重複しないように用語を削除する。

【0045】

第 2 の特徴ベクトル生成部 132 は、検索文の用語と、検索語を含む検索対象文書 22 の用語とを組み合わせ、機械学習で用いられる第 2 の特徴ベクトル 25 を生成する。なお、第 2 の特徴ベクトル生成部 132 は、第 2 の検索部 10B に含まれる。

【0046】

例えば、第 2 の特徴ベクトル生成部 132 は、検索対象文書 22 に含まれる複数の検索対象文から、検索語 28 の表記と一致する用語を含む検索対象文を抽出する。そして、第 2 の特徴ベクトル生成部 132 は、抽出した検索対象文に含まれる用語の意味構造 23 を生成する。用語の意味構造 23 の生成方法は、同義語辞書生成部 11 で用いた方法と同様である。すなわち、第 2 の特徴ベクトル生成部 132 は、検索対象文の意味構造 23 から、検索語 28 の表記と一致する用語の周辺を分離した、検索対象文の用語の意味構造 23 を生成する。そして、第 2 の特徴ベクトル生成部 132 は、生成した検索対象文の用語の意味構造 23 から、特徴ベクトルを生成する。

10

【0047】

また、第 2 の特徴ベクトル生成部 132 は、第 1 の特徴ベクトル生成部 131 によって生成された、検索文の用語の特徴ベクトルを受け取る。

【0048】

また、第 2 の特徴ベクトル生成部 132 は、2 種の特徴ベクトルを比較し、比較結果の特徴ベクトルを生成する。一例として、第 2 の特徴ベクトル生成部 132 は、2 種の特徴ベクトルについて、先頭の素性から順次素性の値を比較し、一致する素性の値を 1 とし、一致しない値を 0 とする比較結果の特徴ベクトルを生成する。

20

【0049】

そして、第 2 の特徴ベクトル生成部 132 は、3 種の特徴ベクトル、すなわち、2 種の意味構造 23 からそれぞれ生成された 2 種の特徴ベクトルと比較結果の特徴ベクトルとを連結して機械学習用の第 2 の特徴ベクトル 25 を生成する。この後、生成された第 2 の特徴ベクトル 25 を、機械学習部 14 が、評価することにより、検索文の用語と、検索対象文の用語とが同義であるか否かを判定することができる。加えて、機械学習部 14 は、検索文の用語と、検索対象文の用語とが表記が同じであっても、意味が同じであるか否かを判定することができる。すなわち、機械学習部 14 は、検索文の用語と、検索対象文の用語とが多義語でないか否かを判定することができる。そして、機械学習部 14 は、判定結果 27 に基づいて、同義語であると判定された用語を含む検索対象文を検索結果として出力する。

30

【0050】

なお、機械学習部 14 により使用される重みベクトル 26 は、以下のように導出される。図 5 は、重みベクトルの導出の一例を示す図である。図 5 に示すように、重みベクトルの導出には、正例の教師例および負例の教師例が使用される。正例の教師例は、同義語であると判定すべき用語と当該用語を含む文章とを併せた情報同士の組である。負例の教師例は、同義語でないと判定すべき用語と当該用語を含む文章とを併せた情報同士の組である。例えば、文章「運用環境のカスタマイズが必要です。」の中の用語「カスタマイズ」と、文章「運用環境の変更により、対処が必要です。」の中の用語「変更」とは、同じ意味を表す異なる表記同士であるため、同義語である。そこで、文章「運用環境のカスタマイズが必要です。」と併せて判定すべき用語「カスタマイズ」と、文章「運用環境の変更により、対処が必要です。」と併せて判定すべき用語「変更」との対は、正例の教師例の一例である。これに対して、文章「運用環境のカスタマイズが必要です。」の中の用語「カスタマイズ」と、文章「組織変更による対象が必要である。」の中の用語「変更」とは、同じ意味を表す異なる表記同士でないため、同義語でない。そこで、文章「運用環境のカスタマイズが必要です。」と併せて判定すべき用語「カスタマイズ」と、文章「組織変

40

50

更による対象が必要である。」と併せて判定すべき用語「変更」との組は、負例の教師例の一例である。

【0051】

図5に示すように、正例の教師例から生成された機械学習用の特徴ベクトルと、負例の教師例から生成された機械学習用の特徴ベクトルとが機械学習部14に入力される。ここでは、正例の教師例として $(x_{12}, 1)$ $(x_{34}, 1)$ $(x_{56}, 1)$ が入力され、負例の教師例として $(x_{17}, -1)$ $(x_{38}, -1)$ $(x_{59}, -1)$ が入力される。例えば、 $(x_{12}, 1)$ の x_{12} は、文書1と文書2とから生成された特徴ベクトルであることを示し、 $(x_{12}, 1)$ の1は、正例であることを示す。 $(x_{17}, -1)$ の x_{17} は、例題の文書1と例題の文書7とから生成された特徴ベクトルであることを示し、 $(x_{17}, -1)$ の-1は、負例であることを示す。

10

【0052】

このような状況の下、機械学習部14は、入力されたそれぞれの機械学習用の特徴ベクトルに対応する重みベクトル26を学習結果として導出する。ここでは、2つの用語に対応する重みベクトル26が表されている。重みベクトル26は、入力された学習用の特徴ベクトルと同じ構造を有する。そして、重みベクトル26には、各素性について、評価用の重みが設定される。この重みベクトル26を用いて、機械学習部14は、第1の特徴ベクトル24を機械学習器で評価し、検索文の用語と、同義語辞書21から抽出された用語とが同義語であるか否かを判定する。この重みベクトル26を用いて、機械学習部14は、第2の特徴ベクトル25を機械学習器で評価し、検索文の用語と、検索対象文書22から抽出された用語とが同義語であるか否かを判定する。

20

【0053】

[第1の検索処理の一例]

図6A~図6Fは、実施例に係る第1の検索処理の一例を示す図である。なお、図6A~図6Fでは、検索文が「画像をクリアに表示するには」である場合について説明する。

【0054】

図6Aに示すように、意味解析部12は、検索文「画像をクリアに表示するには」を受け取ると、受け取った検索文について、意味解析処理により検索文の意味構造23を生成する。ここでは、検索文の意味構造23には、検索文に含まれる用語(「画像」、「クリア」および「表示」)のノードが示される。

30

【0055】

第1の特徴ベクトル生成部131は、検索文に含まれる用語の意味構造23を生成する。以降では、検索文に含まれる用語「画像」、「クリア」および「表示」のうち用語「クリア」についての説明をする。ここでは、検索文に含まれる用語「クリア」は、用語xとする。第1の特徴ベクトル生成部131は、検索文の意味構造23から用語xの周辺を分離して、用語xの意味構造23を生成する。

【0056】

図6Bに示すように、第1の特徴ベクトル生成部131は、用語xの意味構造23から特徴ベクトルを生成する。用語xの特徴ベクトルには、用語xの品詞、意味属性を素性とする特徴情報、用語xと直接アークで接続する用語の品詞、意味属性を素性とする特徴情報が表されている。なお、ここでは、素性として概念記号や方向は、省略している。

40

【0057】

図6Cに示すように、第1の特徴ベクトル生成部131は、同義語辞書21から、検索文の用語xと表記が一致した用語を検索し、検索した用語とペアである他方の用語を抽出する。つまり、同義語辞書21から、検索文の用語xと意味が同じである可能性がある用語が抽出される。抽出される用語は、用語yとする。ここでは、用語yとして、「消去」、「クリア」、「クリア」、「明確」および「鮮明」が抽出される。

【0058】

そして、第1の特徴ベクトル生成部131は、同義語辞書21から、抽出されたそれぞれの用語yに対応付けられた意味構造21cを取得し、取得したそれぞれの用語yの意味

50

構造 2 1 c からそれぞれ特徴ベクトルを生成する。図 6 C の下段 1 段目に、用語 y として抽出された「消去」の特徴ベクトルが表されている。2 段目に、用語 y として抽出された「クリア」の特徴ベクトルが表されている。3 段目に、用語 y として抽出された「クリア」の特徴ベクトルが表されている。4 段目に、用語 y として抽出された「明確」の特徴ベクトルが表されている。5 段目に、用語 y として抽出された「鮮明」の特徴ベクトルが表されている。

【 0 0 5 9 】

図 6 D に示すように、第 1 の特徴ベクトル生成部 1 3 1 は、用語 x の特徴ベクトルと用語 y の特徴ベクトルとを比較し、比較結果の特徴ベクトルを生成する。ここでは、図 6 D の 1 段目に、用語 x の特徴ベクトルと用語 y として抽出された「消去」の特徴ベクトルとの比較結果の特徴ベクトルが表されている。2 段目に、用語 x の特徴ベクトルと用語 y として抽出された「クリア」の特徴ベクトルとの比較結果の特徴ベクトルが表されている。3 段目に、用語 x の特徴ベクトルと用語 y として抽出された「クリア」の特徴ベクトルとの比較結果の特徴ベクトルが表されている。4 段目に、用語 x の特徴ベクトルと用語 y として抽出された「明確」の特徴ベクトルとの比較結果の特徴ベクトルが表されている。5 段目に、用語 x の特徴ベクトルと用語 y として抽出された「鮮明」の特徴ベクトルとの比較結果の特徴ベクトルが表されている。

10

【 0 0 6 0 】

図 6 E に示すように、第 1 の特徴ベクトル生成部 1 3 1 は、用語 x の特徴ベクトルと、用語 y の特徴ベクトルと、用語 x の特徴ベクトルと用語 y の特徴ベクトルとの比較結果の特徴ベクトルとを連結して第 1 の特徴ベクトル 2 4 を生成する。ここでは、図 6 E の 1 行目に、用語 x の特徴ベクトルと用語 y として抽出された「消去」の特徴ベクトルと比較結果の特徴ベクトルとを連結した第 1 の特徴ベクトル 2 4 が表されている。2 行目に、用語 x の特徴ベクトルと用語 y として抽出された「クリア」の特徴ベクトルと比較結果の特徴ベクトルとを連結した第 1 の特徴ベクトル 2 4 が表されている。3 行目に、用語 x の特徴ベクトルと用語 y として抽出された「クリア」の特徴ベクトルと比較結果の特徴ベクトルとを連結した第 1 の特徴ベクトル 2 4 が表されている。4 行目に、用語 x の特徴ベクトルと用語 y として抽出された「明確」の特徴ベクトルと比較結果の特徴ベクトルとを連結した第 1 の特徴ベクトル 2 4 が表されている。5 行目に、用語 x の特徴ベクトルと用語 y として抽出された「鮮明」の特徴ベクトルと比較結果の特徴ベクトルとを連結した第 1 の特徴ベクトル 2 4 が表されている。

20

30

【 0 0 6 1 】

機械学習部 1 4 は、第 1 の特徴ベクトル 2 4 を機械学習器で評価し、用語 x と用語 y とが同義語であるか否かを判定する。ここでは、用語 x である「クリア」と、用語 y として抽出された「クリア」「クリア」「明確」「鮮明」のそれぞれとが、同義語であると判定される。

【 0 0 6 2 】

図 6 F に示すように、検索語生成部 1 5 は、検索文の用語 x である「クリア」を検索語 2 8 とする。そして、検索語生成部 1 5 は、同義語と判定された、用語 y として抽出された「クリア」「クリア」「明確」「鮮明」を検索語 2 8 とする。そして、検索語生成部 1 5 は、検索語 2 8 の中で用語の表記が重複している「クリア」を重複しないように「クリア」を 1 つ残して削除する。これにより、検索文の用語 x である「クリア」に対する検索語は、「クリア」「明確」「鮮明」となる。

40

【 0 0 6 3 】

[第 1 の検索処理の流れ]

つまり、第 1 の検索処理の流れは、以下のようなになる。図 7 は、実施例に係る第 1 の検索処理の流れを示す図である。図 7 に示すように、意味解析部 1 2 は、検索文「画像をクリアに表示するには」を受け取ると、意味構造を持つ同義語辞書 2 1 と付き合わせるために、検索文を形態素解析した後意味解析して意味構造 2 3 を生成する。

【 0 0 6 4 】

50

そして、第 1 の特徴ベクトル生成部 1 3 1 は、用語 x の特徴ベクトルと、用語 y の特徴ベクトルと、用語 x の特徴ベクトルと用語 y の特徴ベクトルとの比較結果の特徴ベクトルとを連結して第 1 の特徴ベクトル 2 4 を生成する。ここでは、用語 x は、検索文の用語の 1 つである「クリア」である。用語 y は、検索文の用語 x と表記が一致した用語とペアである他方の用語であり、「明確」「消去」「鮮明」である。検索文の用語 x である「クリア」の特徴ベクトルが「Z Z 3」であり、用語 y として抽出された「明確」の特徴ベクトルが「Z 2」であるとする。すると、用語 x である「クリア」と用語 y として抽出された「明確」の第 1 の特徴ベクトル 2 4 は、「Z Z 3 __ Z 2 __ Z Z 3 Z 2」となる。「__」は、結合していることを示す。用語 y として抽出された「消去」の特徴ベクトルが「X 2」であるとする。すると、用語 x である「クリア」と用語 y として抽出された「消去」の第 1 の特徴ベクトル 2 4 は、「Z Z 3 __ X 2 __ Z Z 3 X 2」となる。用語 y として抽出された「鮮明」の特徴ベクトルが「Z 6」であるとする。すると、用語 x である「クリア」と用語 y として抽出された「鮮明」の第 1 の特徴ベクトル 2 4 は、「Z Z 3 __ Z 6 __ Z Z 3 Z 6」となる。

10

【 0 0 6 5 】

そして、機械学習部 1 4 は、生成されたそれぞれの第 1 の特徴ベクトル 2 4 を機械学習器で評価し、それぞれの判定結果 2 7 を出力する。ここでは、用語 x である「クリア」と用語 y として抽出された「明確」の第 1 の特徴ベクトル 2 4 を評価した判定結果 2 7 は、同義語であることを示す「」である。用語 x である「クリア」と用語 y として抽出された「消去」の第 1 の特徴ベクトル 2 4 を評価した判定結果 2 7 は、同義語でないことを示す「x」である。用語 x である「クリア」と用語 y として抽出された「鮮明」の第 1 の特徴ベクトル 2 4 を評価した判定結果 2 7 は、同義語であることを示す「」である。これにより、機械学習部 1 4 は、検索文の用語「クリア」と同じ意味の可能性のある「クリア」「消去」「明確」「鮮明」から違う意味の「消去」を対象外とすることができ、同じ意味の同義語だけを抽出できる。

20

【 0 0 6 6 】

そして、検索語生成部 1 5 は、表記の重複を削除し、「クリア」「明確」「鮮明」を検索語として生成する。

【 0 0 6 7 】

[第 2 の検索処理の一例]

図 8 A および図 8 B は、実施例に係る第 2 の検索処理の一例を示す図である。なお、図 8 A および図 8 B では、第 2 の特徴ベクトル生成部 1 3 2 は、検索語生成部 1 5 によって生成された検索語「クリア」「明確」「鮮明」を受け取るものとする。

30

【 0 0 6 8 】

図 8 A に示すように、第 2 の特徴ベクトル生成部 1 3 2 は、検索対象文書 2 2 に含まれる複数の文（検索対象文）から、検索語 2 8 の表記と一致する用語を含む文を抽出する。検索語 2 8 の用語の表記と一致する用語は、用語 y とする。ここでは、検索語 2 8 の「クリア」の表記と一致する用語 y を含む文 1 が抽出される。検索語 2 8 の「クリア」の表記と一致する用語 y を含む文 3 が抽出される。検索語 2 8 の「明確」の表記と一致する用語 y を含む文 5 が抽出される。検索語 2 8 の「クリア」の表記と一致する用語 y を含む文 6 が抽出される。

40

【 0 0 6 9 】

そして、第 2 の特徴ベクトル生成部 1 3 2 は、抽出したそれぞれの文の意味構造 2 8 から、検索語 2 8 の表記と一致するそれぞれの用語 y の意味構造 2 3 を生成する。

【 0 0 7 0 】

図 8 B に示すように、第 2 の特徴ベクトル生成部 1 3 2 は、検索文の用語 x の特徴ベクトルと、用語 y の特徴ベクトルと、用語 x の特徴ベクトルと用語 y の特徴ベクトルとの比較結果の特徴ベクトルとを連結して第 2 の特徴ベクトル 2 5 を生成する。ここでは、図 8 B の 1 行目に、用語 x の特徴ベクトルと文 1 の用語 y 「クリア」の特徴ベクトルと比較結果の特徴ベクトルとを連結した第 2 の特徴ベクトル 2 5 が表されている。2 行目に、用語

50

x の特徴ベクトルと文 3 の用語 y 「クリア」の特徴ベクトルと比較結果の特徴ベクトルとを連結した第 2 の特徴ベクトル 2 5 が表されている。3 行目に、用語 x の特徴ベクトルと文 5 の用語 y 「明確」の特徴ベクトルと比較結果の特徴ベクトルとを連結した第 2 の特徴ベクトル 2 5 が表されている。4 行目に、用語 x の特徴ベクトルと文 6 の用語 y 「クリア」の特徴ベクトルと比較結果の特徴ベクトルとを連結した第 2 の特徴ベクトル 2 5 が表されている。

【 0 0 7 1 】

この後、機械学習部 1 4 は、生成された第 2 の特徴ベクトル 2 5 を機械学習器で評価し、用語 x と用語 y とが同義語であるか否かを判定する。加えて、機械学習部 1 4 は、用語 x と用語 y とが表記が同じであっても、意味が同じであるか否かを判定する。そして、機械学習部 1 4 は、判定結果 2 7 に基づいて、同義語であると判定された用語 y を含む文を検索結果として出力する。

10

【 0 0 7 2 】

[第 2 の検索処理の流れ]

つまり、第 2 の検索処理の流れは、以下のようになる。図 9 は、実施例に係る第 2 の検索処理の流れを示す図である。図 9 に示すように、第 2 の特徴ベクトル生成部 1 3 2 は、検索語 2 8 により検索対象文書 2 2 を検索し、検索語 2 8 の表記と一致する用語を含む文を抽出し、抽出した文の意味構造 2 3 から当該用語の意味構造 2 3 の周辺を分離する。ここでは、文 1 について、文 1 の意味構造 2 3 から該当用語「クリア」の意味構造 2 3 として「p z z 1」が生成される。文 3 について、文 3 の意味構造 2 3 から該当用語「クリア」の意味構造 2 3 として「p x x 1」が生成される。文 5 について、文 5 の意味構造 2 3 から該当用語「明確」の意味構造 2 3 として「p z z 5」が生成される。文 6 について、文 6 の意味構造 2 3 から該当用語「クリア」の意味構造 2 3 として「p w w 6」が生成される。

20

【 0 0 7 3 】

そして、第 2 の特徴ベクトル生成部 1 3 2 は、用語 x の特徴ベクトルと、用語 y の特徴ベクトルと、用語 x の特徴ベクトルと用語 y の特徴ベクトルとの比較結果の特徴ベクトルとを連結して第 2 の特徴ベクトル 2 5 を生成する。ここでは、用語 x は、検索文の用語の 1 つである「クリア」である。用語 y は、検索語 2 8 の表記と一致する文の用語であり、文 1 の「クリア」、文 3 の「クリア」、文 6 の「クリア」および文 5 の「明確」である。検索文の用語 x である「クリア」の特徴ベクトルが「Z Z 3」であり、用語 y である文 1 の「クリア」の特徴ベクトルが「P Z Z 1」であるとする。すると、第 2 の特徴ベクトル 2 5 は、「Z Z 3 _ P Z Z 1 _ Z Z 3 P Z Z 1」となる。「_」は、結合していることを示す。用語 y である文 3 の「クリア」の特徴ベクトルが「P X X 1」であるとする。すると、第 2 の特徴ベクトル 2 5 は、「Z Z 3 _ P X X 1 _ Z Z 3 P X X 1」となる。用語 y である文 6 の「クリア」の特徴ベクトルが「P W W 6」であるとする。すると、第 2 の特徴ベクトル 2 5 は、「Z Z 3 _ P W W 6 _ Z Z 3 P W W 6」となる。用語 y である文 5 の「明確」の特徴ベクトルが「P Z Z 5」であるとする。すると、第 2 の特徴ベクトル 2 5 は、「Z Z 3 _ P Z Z 5 _ Z Z 3 P Z Z 5」となる。

30

【 0 0 7 4 】

そして、機械学習部 1 4 は、生成されたそれぞれの第 2 の特徴ベクトル 2 5 を機械学習器で評価し、それぞれの判定結果 2 7 を出力する。ここでは、検索文の用語「クリア」と文 1 の用語「クリア」の第 2 の特徴ベクトル 2 5 を評価した判定結果 2 7 は、同義語であることを示す「○」である。検索文の用語「クリア」と文 3 の用語「クリア」の第 2 の特徴ベクトル 2 5 を評価した判定結果 2 7 は、同義語でないことを示す「×」である。検索文の用語「クリア」と文 6 の用語「クリア」の第 2 の特徴ベクトル 2 5 を評価した判定結果 2 7 は、同義語でないことを示す「×」である。検索文の用語「クリア」と文 5 の用語「明確」の第 2 の特徴ベクトル 2 5 を評価した判定結果 2 7 は、同義語であることを示す「○」である。

40

【 0 0 7 5 】

50

そして、機械学習部 14 は、判定結果 27 に基づいて、同義語であると判定された用語を含む文を検索結果として出力する。ここでは、文 1 および文 5 が検索結果として出力される。これにより、機械学習部 14 は、検索対象文書 22 内の「クリア」「明確」から、検索語 28 と表記が同じでも意味が違う「クリア」を対象外とすることができ、意味が同じ「クリア」「明確」を抽出できる。

【0076】

[同義語辞書生成処理のフローチャート]

図 10 は、実施例に係る同義語辞書生成処理のフローチャートの一例を示す図である。図 10 に示すように、同義語辞書生成部 11 は、同義語判定結果情報を 1 行ずつ読み込む (ステップ S11)。

10

【0077】

そして、同義語辞書生成部 11 は、同義語判定結果情報のフラグが同義語であることを示す「True」である場合に、用語 1、用語 2 の行番号に基づいて、各用語を含む文章の意味構造 23 を探索し、各用語の意味構造 23 を生成する (ステップ S12)。例えば、同義語辞書生成部 11 は、用語 1 の行番号にある文章の意味構造 23 から用語 1 の周辺を分離した、用語 1 の意味構造 23 を生成する。同義語辞書生成部 11 は、用語 2 の行番号にある文章の意味構造 23 から用語 2 の周辺を分離した、用語 2 の意味構造 23 を生成する。

【0078】

そして、同義語辞書生成部 11 は、用語 1 と用語 2 の区別、各用語の表記、各用語の意味構造 23 を同義語辞書 21 に書き出す (ステップ S13)。例えば、同義語辞書生成部 11 は、フラグ 21a として「用語 1」、用語の表記 21b として用語 1 の表記、意味構造 21c として用語 1 の意味構造をペアの一方として同義語辞書 21 に格納する。同義語辞書生成部 11 は、フラグ 21a として「用語 2」、用語の表記 21b として用語 2 の表記、意味構造 21c として用語 2 の意味構造をペアの他方として同義語辞書 21 に格納する。

20

【0079】

そして、同義語辞書生成部 11 は、同義語判定結果情報の最終行であるか否かを判定する (ステップ S14)。最終行でないと判定した場合には (ステップ S14; No)、同義語辞書生成部 11 は、次の行を読むべく、ステップ S11 に移行する。

30

【0080】

一方、最終行であると判定した場合には (ステップ S14; Yes)、同義語辞書生成部 11 は、同義語辞書 21 の中で、用語 1 と用語 2 の、それぞれの表記と意味構造 23 が全く同じペアを削除する (ステップ S15)。加えて、同義語辞書生成部 11 は、同義語辞書 21 の中で、用語 1 と用語 2 のペアが、他の用語 1 と用語 2 のペアと同じ場合には、重複を削除する (ステップ S16)。そして、同義語辞書生成部 11 は、同義語辞書生成処理を終了する。

【0081】

なお、図 10 のフローチャートで用いた同義語判定結果情報のデータ構造を、図 11 を参照して説明する。図 11 は、同義語判定結果情報のデータ構造の一例を示す図である。図 11 に示すように、同義語判定結果情報は、フラグ、用語 1 の表記、用語 1 を含む文章の意味構造の行番号、用語 2 の表記および用語 2 を含む文章の意味構造の行番号を対応付ける。フラグは、同義語であるか否かを示す。同義語である場合には、「True」が設定され、同義語でない場合には、「False」が設定される。用語 1 の表記は、同義語であるか否かを判定したペアの用語の一方の表記である。用語 1 を含む文章の意味構造の行番号は、用語 1 を含む文章の意味構造がある行番号を示す。用語 2 の表記は、同義語であるか否かを判定したペアの用語の他方の表記である。用語 2 を含む文章の意味構造の行番号は、用語 2 を含む文章の意味構造がある行番号を示す。なお、文章中に、表記の同じ用語が複数ある場合には、同義語判定結果情報のデータ構造に、文章中のカラム位置の情報が付加されても良い。

40

50

【 0 0 8 2 】

[検索処理のフローチャート]

図 1 2 は、実施例に係る検索処理のフローチャートの一例を示す図である。なお、同義語辞書 2 1 は、同義語辞書生成部 1 1 によって生成され、記憶部 2 0 に記憶されているものとする。

【 0 0 8 3 】

図 1 2 に示すように、意味解析部 1 2 は、検索文を受け取ると（ステップ S 2 1 ）、日本語の意味解析処理を実行し、検索文の意味構造 2 3 を生成する（ステップ S 2 2 ）。

【 0 0 8 4 】

続いて、第 1 の特徴ベクトル生成部 1 3 1 は、生成された検索文の意味構造 2 3 と、同義語辞書 2 1 とを用いて、第 1 の特徴ベクトル 2 4 を生成する（ステップ S 2 3 ）。なお、第 1 の特徴ベクトル 2 4 の生成処理のフローチャートは、後述する。

10

【 0 0 8 5 】

そして、機械学習部 1 4 は、生成された第 1 の特徴ベクトル 2 4 の評価をすべく、機械学習を実行する（ステップ S 2 4 ）。機械学習部 1 4 は、予め、教師例に基づいて機械学習し、学習結果として重みベクトル 2 6 を導出するものとする。そして、機械学習部 1 4 は、学習結果である重みベクトル 2 6 を用いて、機械学習を実行する。なお、機械学習の実行処理のフローチャートは、後述する。

【 0 0 8 6 】

そして、検索語生成部 1 5 は、機械学習の判定結果 2 7 に基づいて、検索語 2 8 を生成する（ステップ S 2 5 ）。なお、検索語 2 8 の生成処理のフローチャートは、後述する。

20

【 0 0 8 7 】

続いて、第 2 の特徴ベクトル生成部 1 3 2 は、生成された検索語 2 8 と、検索対象文書 2 2 とを用いて、第 2 の特徴ベクトル 2 5 を生成する（ステップ S 2 6 ）。なお、第 2 の特徴ベクトル 2 5 の生成処理のフローチャートは、後述する。

【 0 0 8 8 】

そして、機械学習部 1 4 は、生成された第 2 の特徴ベクトル 2 5 の評価をすべく、機械学習を実行する（ステップ S 2 7 ）。機械学習部 1 4 は、ステップ S 2 4 で用いた重みベクトル 2 6 を用いて、機械学習を実行する。なお、機械学習の実行処理のフローチャートは、後述する。

30

【 0 0 8 9 】

そして、機械学習部 1 4 は、機械学習の判定結果 2 7 に基づいて、検索文に対応する検索対象文を検索結果として出力する（ステップ S 2 8 ）。そして、制御部 1 0 は、検索処理を終了する。

【 0 0 9 0 】

[第 1 の特徴ベクトル生成処理のフローチャート]

図 1 3 は、実施例に係る第 1 の特徴ベクトル生成処理のフローチャートの一例を示す図である。

【 0 0 9 1 】

第 1 の特徴ベクトル生成部 1 3 1 は、検索文の意味構造 2 3 を受け取る（ステップ S 3 1 ）。第 1 の特徴ベクトル生成部 1 3 1 は、検索文の意味構造 2 3 の用語の表記から、検索文の用語を特定する（ステップ S 3 2 ）。第 1 の特徴ベクトル生成部 1 3 1 は、検索文の用語（用語 x ）の意味構造 2 3 から、用語 x の特徴ベクトルを生成する（ステップ S 3 3 ）。例えば、第 1 の特徴ベクトル生成部 1 3 1 は、検索文の意味構造 2 3 から用語 x の周辺を分離した、用語 x の意味構造 2 3 を生成し、生成した用語 x の意味構造 2 3 から、用語 x の特徴ベクトルを生成する。

40

【 0 0 9 2 】

続いて、第 1 の特徴ベクトル生成部 1 3 1 は、同義語辞書 2 1 から、検索文の用語 x と表記が一致する用語を検索する（ステップ S 3 4 ）。そして、第 1 の特徴ベクトル生成部 1 3 1 は、検索した用語とペアである他方の用語（用語 y ）を抽出し、用語 y に対応付け

50

られた意味構造 2 1 c から、用語 y の特徴ベクトルを生成する（ステップ S 3 5）。

【 0 0 9 3 】

続いて、第 1 の特徴ベクトル生成部 1 3 1 は、検索文の用語 x の特徴ベクトルと用語 y の特徴ベクトルとを比較し、比較結果の特徴ベクトルを生成する（ステップ S 3 6）。そして、第 1 の特徴ベクトル生成部 1 3 1 は、検索文の用語 x、用語 y、比較結果の特徴ベクトルを連結して、第 1 の特徴ベクトル 2 4 を生成する（ステップ S 3 7）。

【 0 0 9 4 】

そして、第 1 の特徴ベクトル生成部 1 3 1 は、未処理の検索文の用語は有るか否かを判定する（ステップ S 3 8）。未処理の検索文の用語は有ると判定した場合には（ステップ S 3 8 ; Y e s ）、第 1 の特徴ベクトル生成部 1 3 1 は、次の検索文の用語を特定すべく、ステップ S 3 2 に移行する。

【 0 0 9 5 】

一方、未処理の検索文の用語は無いと判定した場合には（ステップ S 3 8 ; N o ）、第 1 の特徴ベクトル生成部 1 3 1 は、第 1 の特徴ベクトル生成処理を終了する。

【 0 0 9 6 】

[第 2 の特徴ベクトル生成処理のフローチャート]

図 1 4 は、実施例に係る第 2 の特徴ベクトル生成処理のフローチャートの一例を示す図である。

【 0 0 9 7 】

第 2 の特徴ベクトル生成部 1 3 2 は、検索文の用語に対する検索語 2 8 を受け取る（ステップ S 4 1）。第 2 の特徴ベクトル生成部 1 3 2 は、検索語 2 8 をキーとして、検索対象文書 2 2 の意味構造 2 3 の表記を検索する（ステップ S 4 2）。第 2 の特徴ベクトル生成部 1 3 2 は、検索語 2 8 と表記が一致した用語について、検索対象文書 2 2 の意味構造 2 3 から、用語の意味構造 2 3 を生成する（ステップ S 4 3）。

【 0 0 9 8 】

第 2 の特徴ベクトル生成部 1 3 2 は、検索語 2 8 と表記が一致した用語、検索対象文書 2 2 中の当該用語の位置情報および当該用語の意味構造 2 3 を対応付けて、一時的に保持する（ステップ S 4 4）。

【 0 0 9 9 】

そして、第 2 の特徴ベクトル生成部 1 3 2 は、検索文の用語（用語 x）の特徴ベクトルを取得する（ステップ S 4 5）。例えば、第 2 の特徴ベクトル生成部 1 3 2 は、第 1 の特徴ベクトル生成部 1 3 1 によって生成された用語 x の特徴ベクトルを取得すれば良い。

【 0 1 0 0 】

そして、第 2 の特徴ベクトル生成部 1 3 2 は、検索対象文書 2 2 中で、検索語 2 8 と表記が一致した用語（用語 y）の意味構造 2 3 から特徴ベクトルを生成する（ステップ S 4 6）。例えば、第 2 の特徴ベクトル生成部 1 3 2 は、一時的に保持した、検索語 2 8 と表記が一致した用語に対応付けられた当該用語の意味構造 2 3 から特徴ベクトルを生成する。

【 0 1 0 1 】

続いて、第 2 の特徴ベクトル生成部 1 3 2 は、検索文の用語 x の特徴ベクトルと用語 y の特徴ベクトルとを比較し、比較結果の特徴ベクトルを生成する（ステップ S 4 7）。そして、第 2 の特徴ベクトル生成部 1 3 2 は、検索文の用語 x、用語 y、比較結果の特徴ベクトルを連結して、第 2 の特徴ベクトル 2 5 を生成する（ステップ S 4 8）。

【 0 1 0 2 】

そして、第 2 の特徴ベクトル生成部 1 3 2 は、未処理の検索文の用語は有るか否かを判定する（ステップ S 4 9）。未処理の検索文の用語は有ると判定した場合には（ステップ S 4 9 ; Y e s ）、第 2 の特徴ベクトル生成部 1 3 2 は、次の検索文の用語を処理すべく、ステップ S 4 2 に移行する。

【 0 1 0 3 】

一方、未処理の検索文の用語は無いと判定した場合には（ステップ S 4 9 ; N o ）、第

10

20

30

40

50

2 の特徴ベクトル生成部 1 3 2 は、第 2 の特徴ベクトル生成処理を終了する。

【 0 1 0 4 】

[機械学習処理のフローチャート]

図 1 5 は、実施例に係る機械学習処理のフローチャートの一例を示す図である。

【 0 1 0 5 】

図 1 5 に示すように、機械学習部 1 4 は、第 1 または第 2 の特徴ベクトル 2 5 , 2 5 を受け取る (ステップ S 5 1)。例えば、機械学習部 1 4 は、第 1 の特徴ベクトル生成部 1 3 1 から第 1 の特徴ベクトル 2 4 を受け取る。機械学習部 1 4 は、第 2 の特徴ベクトル生成部 1 3 2 から第 2 の特徴ベクトル 2 5 を受け取る。

【 0 1 0 6 】

機械学習部 1 4 は、受け取った第 1 の特徴ベクトル 2 4 または第 2 の特徴ベクトル 2 5 を機械学習器に入力し、機械学習器による評価を行う (ステップ S 5 2)。例えば、機械学習部 1 4 は、第 1 の特徴ベクトル 2 4 を受け取った場合には、第 1 の特徴ベクトル 2 4 と重みベクトル 2 6 との内積を計算し、総合評価値を導出する。そして、機械学習部 1 4 は、導出した総合評価値を予め設定された閾値で判定することにより、同義語であるか否かの判定結果 2 7 を決定し、決定した判定結果 2 7 を出力する (ステップ S 5 3)。また、機械学習部 1 4 は、第 2 の特徴ベクトル 2 5 を受け取った場合にも、第 1 の特徴ベクトル 2 4 を受け取った場合と同様に、重みベクトル 2 6 および予め設定された閾値を用いて、判定結果 2 7 を決定し、決定した判定結果 2 7 を出力する (ステップ S 5 3)。

【 0 1 0 7 】

機械学習部 1 4 は、受け取った特徴ベクトルが第 1 の特徴ベクトル 2 4 であるか否かを判定する (ステップ S 5 4)。受け取った特徴ベクトルが第 1 の特徴ベクトル 2 4 であると判定した場合には (ステップ S 5 4 ; Y e s)、機械学習部 1 4 は、検索語 2 8 の生成処理を行う (ステップ S 5 5)。なお、検索語生成処理のフローチャートは、後述する。そして、機械学習部 1 4 は、機械学習処理を終了する。

【 0 1 0 8 】

一方、受け取った特徴ベクトルが第 1 の特徴ベクトル 2 4 でないと判定した場合には (ステップ S 5 4 ; N o)、機械学習部 1 4 は、同義語であると判定された用語について、その位置情報から検索対象文書 2 2 内の検索対象文を取得する (ステップ S 5 6)。そして、機械学習部 1 4 は、検索対象文を検索結果として出力する (ステップ S 5 7)。そして、機械学習部 1 4 は、機械学習処理を終了する。

【 0 1 0 9 】

[検索語生成処理のフローチャート]

図 1 6 は、実施例に係る検索語生成処理のフローチャートの一例を示す図である。

【 0 1 1 0 】

図 1 6 に示すように、検索語生成部 1 5 は、機械学習部 1 4 から判定結果 2 7 を受け取る (ステップ S 6 1)。検索語生成部 1 5 は、判定結果 2 7 から、同義語として判定された用語 x と用語 y を読み込む (ステップ S 6 2)。そして、検索語生成部 1 5 は、用語 x に対する用語 y の表記の重複を削除する (ステップ S 6 3)。

【 0 1 1 1 】

検索語生成部 1 5 は、検索文の用語 x に対する用語 y を検索語として出力する (ステップ S 6 4)。そして、検索語生成部 1 5 は、未処理の検索文の用語は有るか否かを判定する (ステップ S 6 5)。未処理の検索文の用語は有ると判定した場合には (ステップ S 6 5 ; Y e s)、検索語生成部 1 5 は、次の検索文の用語に対する判定結果 2 7 を処理すべく、ステップ S 6 2 に移行する。

【 0 1 1 2 】

一方、未処理の検索文の用語は無いと判定した場合には (ステップ S 6 5 ; N o)、検索語生成部 1 5 は、検索語生成処理を終了する。

【 0 1 1 3 】

[実施例の効果]

10

20

30

40

50

このようにして、情報処理装置 1 は、検索対象文書 2 2 から特定の文章を検索する場合に、検索文を受け付ける。情報処理装置 1 は、受け付けた検索文を意味解析して、受け付けた検索文に含まれる単語の意味を示す意味構造 2 3 を生成する。情報処理装置 1 は、単語と当該単語の意味を示す意味構造 2 3 とを対応付けて記憶する同義語辞書 2 1 を参照して、生成した意味構造 2 3 に対応付けられた単語を特定する。情報処理装置 1 は、特定した単語が検索対象文書 2 2 に含まれるか否かの判定を行い、判定結果に応じた情報を出力する。かかる構成によれば、情報処理装置 1 は、検索文の単語の意味と、同義語辞書 2 1 の単語の意味とを組み合わせ得られる単語を特定することで、検索対象文書 2 2 から特定の文書を検索する際の検索語を特定することができる。この結果、情報処理装置 1 は、検索語を用いて、検索対象文書 2 2 から特定の文章を検索すると、検索結果に含まれるノイズを低減することが可能となる。すなわち、情報処理装置 1 は、検索対象文書 2 2 から意図しない特定の文章を検索することを低減することが可能となる。

10

20

30

40

50

【0114】

また、情報処理装置 1 は、特定した単語が検索対象文書 2 2 に含まれると判定した場合には、以下の処理を行う。情報処理装置 1 は、検索対象文書 2 2 に含まれる複数の単語それぞれを該複数の単語それぞれの意味を示す意味構造 2 3 に対応付けて記憶する第 2 の記憶部を参照して、検索文に含まれる単語の意味構造 2 3 に対応付けられた単語を特定する。情報処理装置 1 は、該特定した単語が検索対象文書 2 2 の中のいずれの文章に含まれるか否かの判定を行い、判定結果に応じた情報を出力する。かかる構成によれば、情報処理装置 1 は、検索文の単語の意味と、検索対象文書 2 2 に含まれる、特定した単語（検索語 2 8）の意味とを組み合わせ、検索対象文書 2 2 から特定の文章を検索することで、検索結果に含まれるノイズを低減することができる。例えば、情報処理装置 1 は、表記が同じで意味が異なる多義語を持つ特定の文章を排除することができる。

【0115】

また、情報処理装置 1 は、第 2 の記憶部を参照して、同義語辞書 2 1 を参照して特定した単語と一致する単語に対応付けられた意味構造 2 3 を特定する。情報処理装置 1 は、特定した意味構造 2 3 と、検索文に含まれる単語の意味構造 2 3 とを用いて、検索文に含まれる単語の意味構造 2 3 に対応付けられた単語を特定する。かかる構成によれば、情報処理装置 1 は、検索文の単語の意味と、検索対象文書 2 2 に含まれる、特定した単語（検索語 2 8）の意味とを組み合わせ得られる単語を特定することで、検索対象文書 2 2 内の特定の文章に含まれる単語を特定することができる。この結果、情報処理装置 1 は、特定した単語を用いて、検索対象文書 2 2 内の特定の文章を検索でき、特定の文章がノイズとなることを低減することができる。

【0116】

また、情報処理装置 1 は、受け付けた検索文の意味構造 2 3 から、検索文に含まれる単語と、当該単語と直接関係する他の単語との関係を示す情報を抽出する。情報処理装置 1 は、抽出した情報を意味構造 2 3 として生成する。かかる構成によれば、情報処理装置 1 は、検索文に含まれる単語と直接関係のある他の単語との関係を示す情報を単語の意味構造 2 3 として生成することで、検索文に含まれる単語の意味を示す情報を生成できる。

【0117】**[その他]**

なお、実施例に係る同義語辞書 2 1 は、同義語として判定されたそれぞれの用語（用語 1、用語 2）を、それぞれの用語の意味構造 2 3 と対応付けて記憶すると説明した。しかしながら、同義語辞書 2 1 は、これに限定されず、同義語として判定されたそれぞれの用語をそれぞれの用語の特徴ベクトルと対応付けて記憶しても良い。かかる場合には、同義語辞書生成部 1 1 は、同義語と判定された各用語が出現する文章について、意味解析処理により、それぞれの意味構造 2 3 を生成する。同義語辞書生成部 1 1 は、用語 1 が出現する文章の意味構造 2 3 から用語 1 の周辺を分離した、用語 1 の意味構造 2 3 を生成する。同義語辞書生成部 1 1 は、用語 2 が出現する文章の意味構造 2 3 から用語 2 の周辺を分離した、用語 2 の意味構造 2 3 を生成する。そして、同義語辞書生成部 1 1 は、生成した用

語 1 の意味構造 2 3 から用語 1 の特徴ベクトルを生成する。同義語辞書生成部 1 1 は、生成した用語 2 の意味構造 2 3 から用語 2 の特徴ベクトルを生成する。そして、同義語辞書生成部 1 1 は、用語 1 と用語 2 について、用語の表記、用語の特徴ベクトルを同義語辞書 2 1 に書き出せば良い。これにより、第 1 の特徴ベクトル生成部 1 3 1 は、同義語辞書 2 1 から、該当する用語の意味構造 2 3 の代わりに該当する用語の特徴ベクトルを直接取得することができる。

【 0 1 1 8 】

また、図示した情報処理装置 1 の各構成要素は、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、情報処理装置 1 の分散・統合の具体的態様は図示のものに限られず、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。例えば、意味解析部 1 2 と、第 1 の特徴ベクトル生成部 1 3 1 とを 1 つの部として統合しても良い。また、機械学習部 1 4 を、重みベクトル 2 6 を学習する第 1 の機械学習部と、第 1 の特徴ベクトル 2 4 を評価する第 2 の機械学習部と、第 2 の特徴ベクトル 2 5 を評価する第 3 の機械学習部とに分散しても良い。また、記憶部 2 0 を情報処理装置 1 の外部装置としてネットワーク経由で接続するようにしても良い。

10

【 0 1 1 9 】

また、上記実施例で説明した各種の処理は、予め用意されたプログラムをパーソナルコンピュータやワークステーションなどのコンピュータで実行することによって実現することができる。そこで、以下では、図 1 に示した情報処理装置 1 と同様の機能を実現する検索プログラムを実行するコンピュータの一例を説明する。図 1 7 は、検索プログラムを実行するコンピュータの一例を示す図である。

20

【 0 1 2 0 】

図 1 7 に示すように、コンピュータ 2 0 0 は、各種演算処理を実行する CPU 2 0 3 と、ユーザからのデータの受け付ける入力装置 2 1 5 と、表示装置 2 0 9 を制御する表示制御部 2 0 7 とを有する。また、コンピュータ 2 0 0 は、記憶媒体からプログラムなどを読み取るドライブ装置 2 1 3 と、ネットワークを介して他のコンピュータとの間でデータの授受を行う通信制御部 2 1 7 とを有する。また、コンピュータ 2 0 0 は、各種情報を一時記憶するメモリ 2 0 1 と、HDD 2 0 5 を有する。そして、メモリ 2 0 1、CPU 2 0 3、HDD 2 0 5、表示制御部 2 0 7、ドライブ装置 2 1 3、入力装置 2 1 5、通信制御部 2 1 7 は、バス 2 1 9 で接続されている。

30

【 0 1 2 1 】

ドライブ装置 2 1 3 は、例えばリムーバブルディスク 2 1 1 用の装置である。HDD 2 0 5 は、検索プログラム 2 0 5 a および検索処理関連情報 2 0 5 b を記憶する。

【 0 1 2 2 】

CPU 2 0 3 は、検索プログラム 2 0 5 a を読み出して、メモリ 2 0 1 に展開し、プロセスとして実行する。かかるプロセスは、情報処理装置 1 の各機能部に対応する。検索処理関連情報 2 0 5 b は、例えば、同義語辞書 2 1、検索対象文書 2 2、意味構造 2 3、第 1 の特徴ベクトル 2 4、第 2 の特徴ベクトル 2 5、重みベクトル 2 6、判定結果 2 7 および検索語 2 8 に対応する。そして、例えばリムーバブルディスク 2 1 1 が、検索プログラム 2 0 5 a などの各情報を記憶する。

40

【 0 1 2 3 】

なお、検索プログラム 2 0 5 a については、必ずしも最初から HDD 2 0 5 に記憶させておかなくても良い。例えば、コンピュータ 2 0 0 に挿入されるフレキシブルディスク (FD)、CD-ROM、DVD ディスク、光磁気ディスク、IC カードなどの「可搬用の物理媒体」に当該プログラムを記憶させておく。そして、コンピュータ 2 0 0 がこれらから検索プログラム 2 0 5 a を読み出して実行するようにしても良い。

【 符号の説明 】

【 0 1 2 4 】

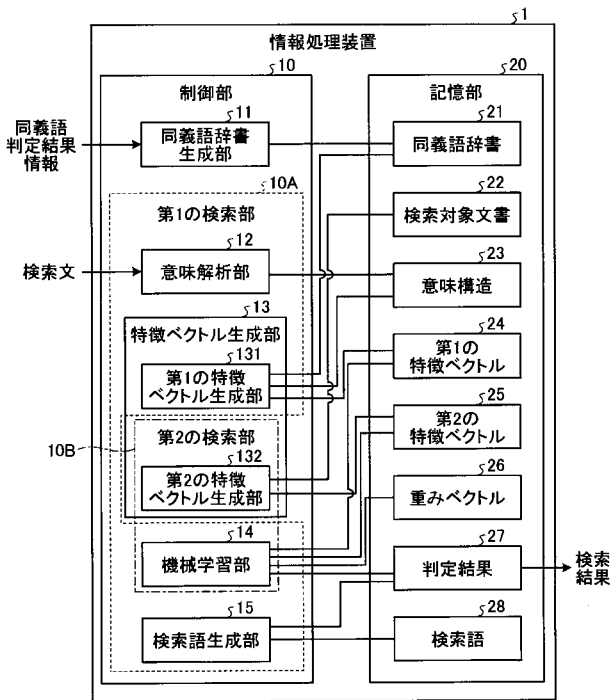
1 情報処理装置

50

- 1 0 制御部
- 1 0 A 第 1 の 検 索 部
- 1 0 B 第 2 の 検 索 部
- 1 1 同 義 語 辞 書 生 成 部
- 1 2 意 味 解 析 部
- 1 3 特 徴 ベ ク ト ル 生 成 部
- 1 3 1 第 1 の 特 徴 ベ ク ト ル 生 成 部
- 1 3 2 第 2 の 特 徴 ベ ク ト ル 生 成 部
- 1 4 機 械 学 習 部
- 1 5 検 索 語 生 成 部
- 2 0 記 憶 部
- 2 1 同 義 語 辞 書
- 2 2 検 索 対 象 文 書
- 2 3 意 味 構 造
- 2 4 第 1 の 特 徴 ベ ク ト ル
- 2 5 第 2 の 特 徴 ベ ク ト ル
- 2 6 重 み ベ ク ト ル
- 2 7 判 定 結 果
- 2 8 検 索 語

【 図 1 】

実施例に係る情報処理装置の構成の一例を示す機能ブロック図



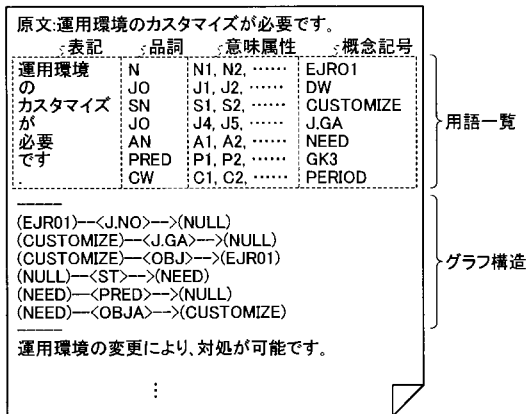
【 図 2 】

実施例に係る同義語辞書の一例を示す図

	21a	21b	21c
	フラグ	用語の表記	意味構造
同義語	用語1	一意	z1
	用語2	ユニーク	z2
同義語	用語1	クリア	x1
	用語2	クリア	x2
	⋮	⋮	⋮

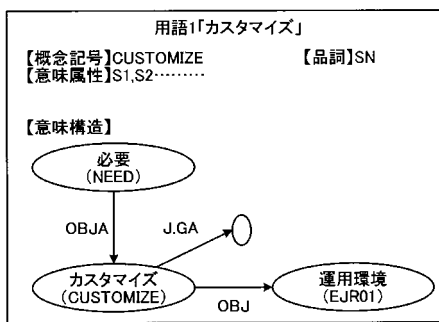
【 図 3 A 】

意味構造の一例を示す図(1)



【 図 3 C 】

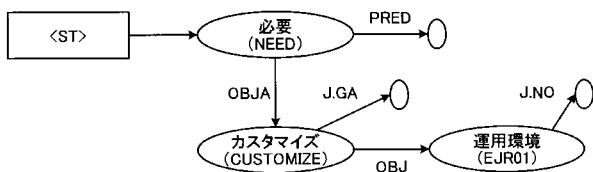
用語の意味構造の一例を示す図



【 図 3 B 】

意味構造の一例を示す図(2)

原文:運用環境のカスタマイズが必要です。



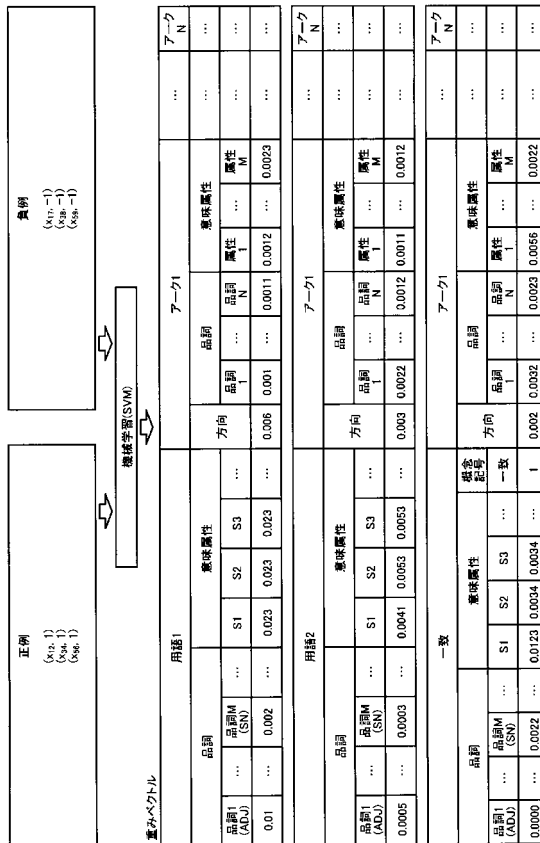
【 図 4 】

実施例に係る特徴ベクトルの構成の一例を示す図

対象ノードの情報 10				関連ノードの情報 10			
用語				アーク1			
品詞		意味属性		品詞		意味属性	
品詞 (ADJ)	品詞 (SN)	S1	S2	S3	品詞 1	品詞 SN	属性 1
0	...	1	1	0	0	0	0
...
...
...

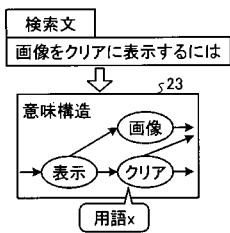
【 図 5 】

重みベクトルの導出の一例を示す図



【図6A】

実施例に係る第1の検索処理の一例を示す図(1)



【図6B】

実施例に係る第1の検索処理の一例を示す図(2)

用語x(検索文)の特徴ベクトル											
用語x(検索文)						アーク1					
品詞			意味属性			品詞			意味属性		
品詞1 (ADJ)	品詞7 (SN)	品詞10	ANSN V&N	...	NSN	品詞1	品詞30	ANSN V&N	...	NSN	...
0	1	0	1	...	1	...	0	0	...	0	...

【図6C】

実施例に係る第1の検索処理の一例を示す図(3)

フラグ	用語の表記	意味構造	用語y
用語1	クリア	クリアの意味構造1	「用語y」
用語2	消去	消去の意味構造1	「用語y」
用語1	クリア	クリアの意味構造1	「用語y」
用語2	クリア	クリアの意味構造2	「用語y」
用語1	クリア	クリアの意味構造2	「用語y」
用語2	クリア	クリアの意味構造3	「用語y」
...
用語1	クリア	クリアの意味構造3	「用語y」
用語2	明確	明確の意味構造1	「用語y」
用語1	鮮明	鮮明の意味構造1	「用語y」
用語2	クリア	クリアの意味構造3	「用語y」

用語y(消去)の特徴ベクトル											
用語y(同義語辞書の用語)						アーク1					
品詞1 (ADJ)	品詞7 (SN)	品詞10	ANSN V&N	...	NSN	品詞1	品詞30	ANSN V&N	...	NSN	...
1	...	0	0	1	...	1	...	0	0	...	0

用語y(クリア)の特徴ベクトル											
用語y(同義語辞書の用語)						アーク1					
品詞1 (ADJ)	品詞7 (SN)	品詞10	ANSN V&N	...	NSN	品詞1	品詞30	ANSN V&N	...	NSN	...
1	...	0	0	1	...	1	...	0	0	...	0

用語y(クリア)の特徴ベクトル											
用語y(同義語辞書の用語)						アーク1					
品詞1 (ADJ)	品詞7 (SN)	品詞10	ANSN V&N	...	NSN	品詞1	品詞30	ANSN V&N	...	NSN	...
1	...	0	0	1	...	1	...	0	0	...	0

用語y(明確)の特徴ベクトル											
用語y(同義語辞書の用語)						アーク1					
品詞1 (ADJ)	品詞7 (SN)	品詞10	ANSN V&N	...	NSN	品詞1	品詞30	ANSN V&N	...	NSN	...
0	...	1	0	1	...	1	...	0	0	...	0

用語y(鮮明)の特徴ベクトル											
用語y(同義語辞書の用語)						アーク1					
品詞1 (ADJ)	品詞7 (SN)	品詞10	ANSN V&N	...	NSN	品詞1	品詞30	ANSN V&N	...	NSN	...
0	...	1	0	1	...	1	...	0	0	...	0

【図6D】

実施例に係る第1の検索処理の一例を示す図(4)

用語x-用語y(消去)の比較結果の特徴ベクトル											
用語y(同義語辞書の用語)						アーク1					
品詞1 (ADJ)	品詞7 (SN)	品詞10	ANSN V&N	...	NSN	品詞1	品詞30	ANSN V&N	...	NSN	...
1	...	0	0	1	...	1	...	0	0	...	0

用語x-用語y(クリア)の比較結果の特徴ベクトル											
用語y(同義語辞書の用語)						アーク1					
品詞1 (ADJ)	品詞7 (SN)	品詞10	ANSN V&N	...	NSN	品詞1	品詞30	ANSN V&N	...	NSN	...
1	...	0	0	1	...	1	...	0	0	...	0

用語x-用語y(クリア)の比較結果の特徴ベクトル											
用語y(同義語辞書の用語)						アーク1					
品詞1 (ADJ)	品詞7 (SN)	品詞10	ANSN V&N	...	NSN	品詞1	品詞30	ANSN V&N	...	NSN	...
1	...	0	0	1	...	1	...	0	0	...	0

用語x-用語y(明確)の比較結果の特徴ベクトル											
用語y(同義語辞書の用語)						アーク1					
品詞1 (ADJ)	品詞7 (SN)	品詞10	ANSN V&N	...	NSN	品詞1	品詞30	ANSN V&N	...	NSN	...
0	...	1	0	1	...	1	...	0	0	...	0

用語x-用語y(鮮明)の比較結果の特徴ベクトル											
用語y(同義語辞書の用語)						アーク1					
品詞1 (ADJ)	品詞7 (SN)	品詞10	ANSN V&N	...	NSN	品詞1	品詞30	ANSN V&N	...	NSN	...
0	...	1	0	1	...	1	...	0	0	...	0

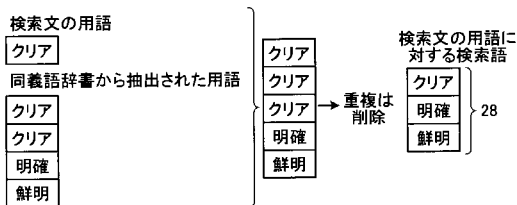
【図 6 E】

実施例に係る第1の検索処理の一例を示す図(5)

用語xの特徴ベクトル (検索文の用語)	用語yの特徴ベクトル (同義語辞書の用語)	比較結果の 特徴ベクトル
クリア	消去	クリア-消去の比較
クリア	クリア	クリア-クリアの比較
クリア	クリア	クリア-クリアの比較
クリア	明確	クリア-明確の比較
クリア	鮮明	クリア-鮮明の比較

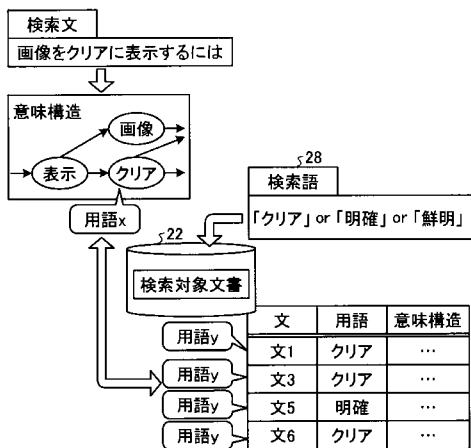
【図 6 F】

実施例に係る第1の検索処理の一例を示す図(6)



【図 8 A】

実施例に係る第2の検索処理の一例を示す図(1)



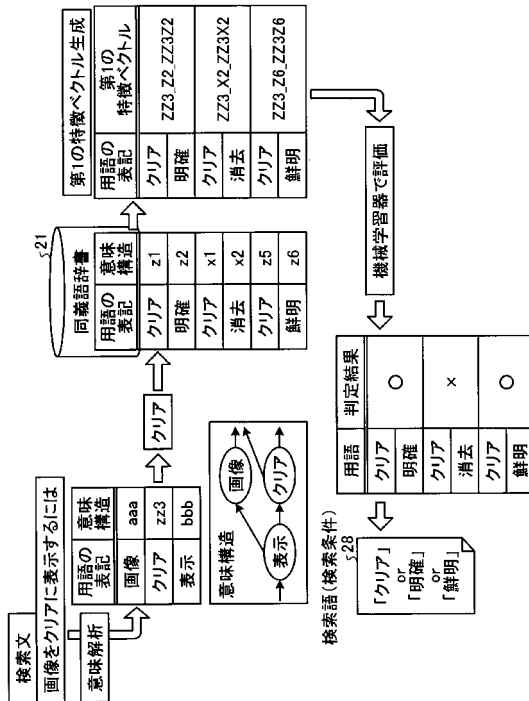
【図 8 B】

実施例に係る第2の検索処理の一例を示す図(2)

用語xの特徴ベクトル (検索文の用語)	用語yの特徴ベクトル (検索対象文書から検索した、 検索語と表記が一致した用語)	比較結果の 特徴ベクトル
クリア	文1のクリア	クリア-文1のクリアの比較
クリア	文3のクリア	クリア-文3のクリアの比較
クリア	文5の明確	クリア-文5の明確の比較
クリア	文6のクリア	クリア-文6のクリアの比較

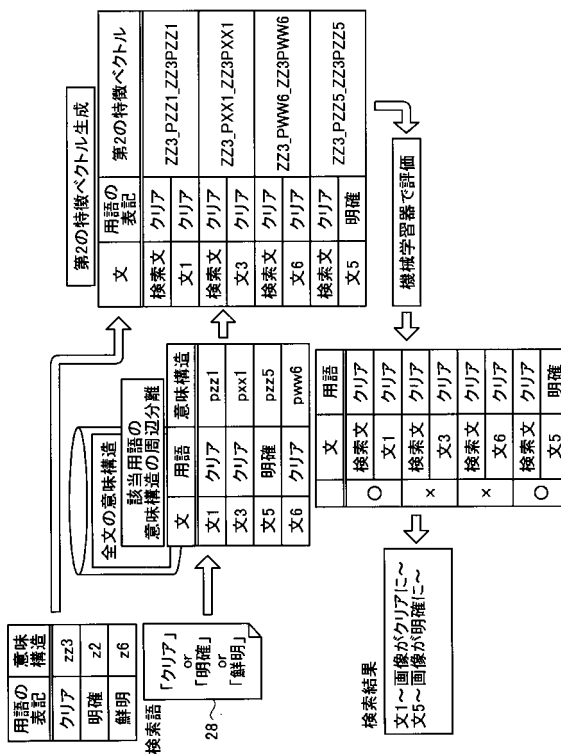
【図 7】

実施例に係る第1の検索処理の流れを示す図



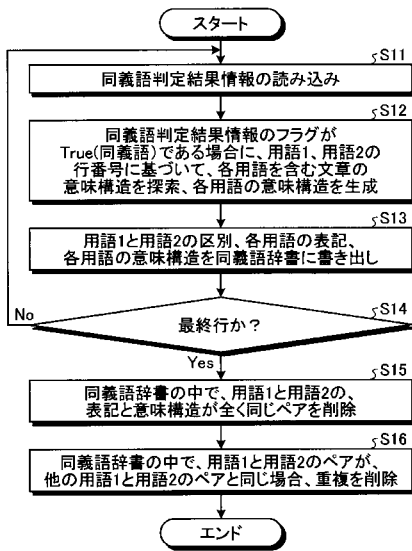
【図 9】

実施例に係る第2の検索処理の流れを示す図



【 図 1 0 】

実施例に係る同義語辞書生成処理のフローチャートの一例を示す図



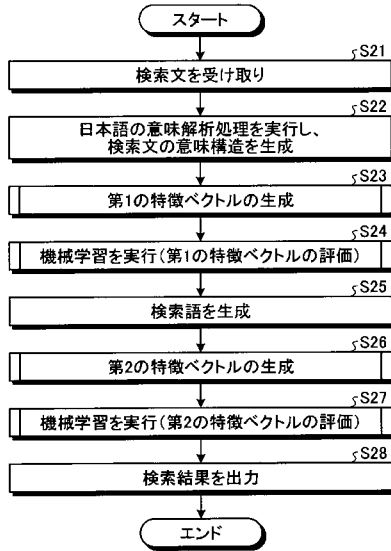
【 図 1 1 】

同義語判定結果情報のデータ構造の一例を示す図

フラグ	用語1の表記	用語1を含む文章の意味構造の行番号	用語2の表記	用語2を含む文章の意味構造の行番号
True	一意	249582	ユニーク	777752
False	クリア	1106786	クリア	35216
True	資料	1402	コンテンツ	33110
True	文字	150	単語	1075

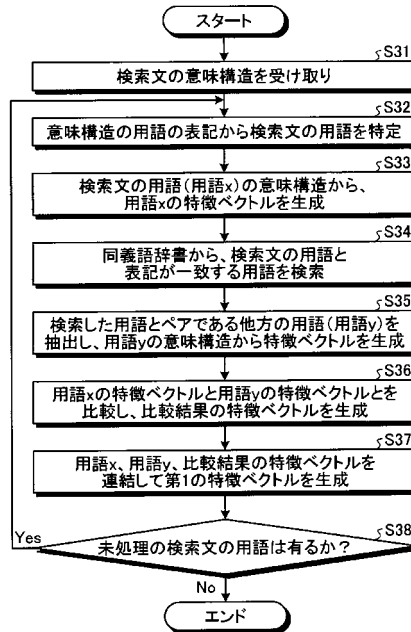
【 図 1 2 】

実施例に係る検索処理のフローチャートの一例を示す図



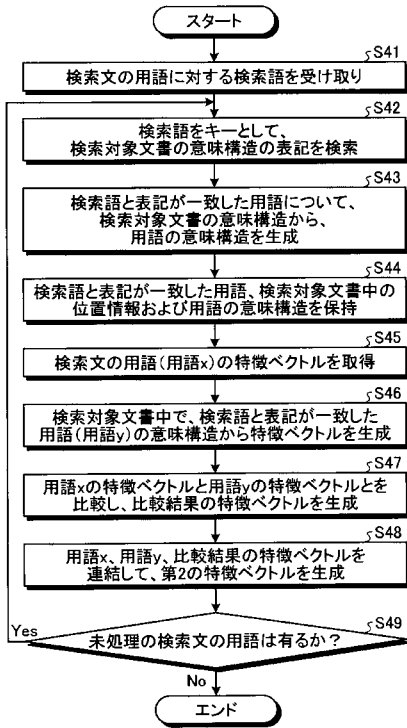
【 図 1 3 】

実施例に係る第1の特徴ベクトル生成処理のフローチャートの一例を示す図



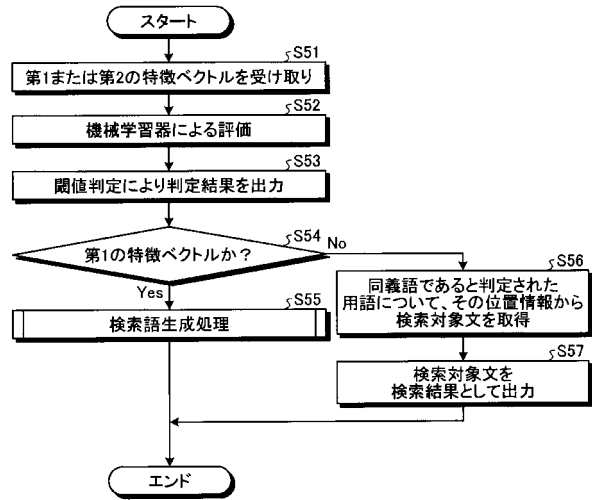
【 図 1 4 】

実施例に係る第2の特徴ベクトル生成処理のフローチャートの一例を示す図



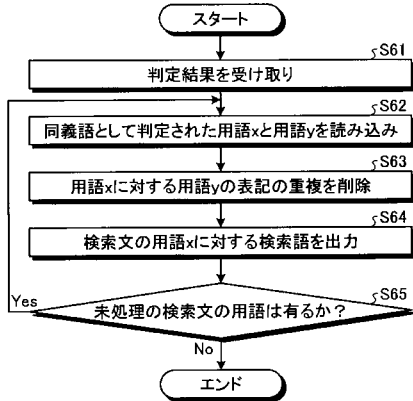
【 図 1 5 】

実施例に係る機械学習処理のフローチャートの一例を示す図



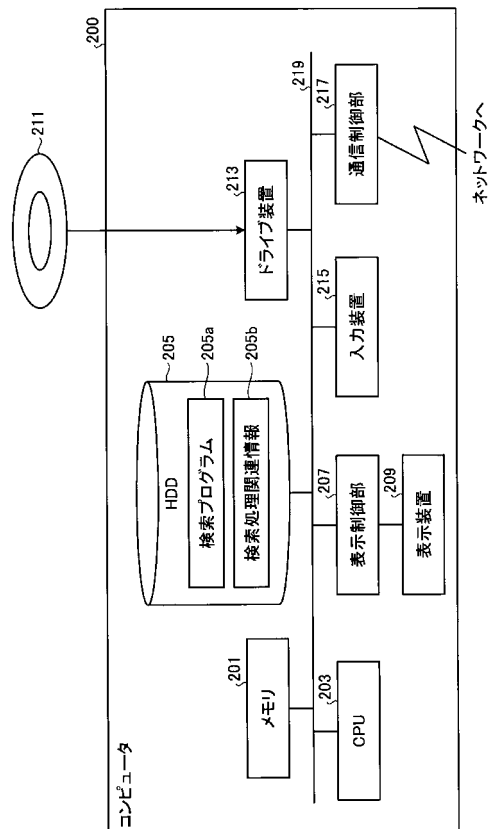
【 図 1 6 】

実施例に係る検索語生成処理のフローチャートの一例を示す図



【 図 1 7 】

検索プログラムを実行するコンピュータの一例を示す図



【 図 1 8 】

検索手法の一例を示す図

