

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2024年1月18日(18.01.2024)



(10) 国際公開番号

WO 2024/013991 A1

(51) 国際特許分類:
G06F 16/35 (2019.01)

(21) 国際出願番号: PCT/JP2022/027902

(22) 国際出願日: 2022年7月15日(15.07.2022)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人: 富士通株式会社(FUJITSU LIMITED)
[JP/JP]; 〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 Kanagawa (JP).

(72) 発明者: 片岡 正弘 (KATAOKA, Masahiro);
〒2118588 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 Kanagawa (JP). 岩田 彩 (IWATA, Aya); 〒2118588 神奈川県川崎市中原区上小田中4丁目1番1

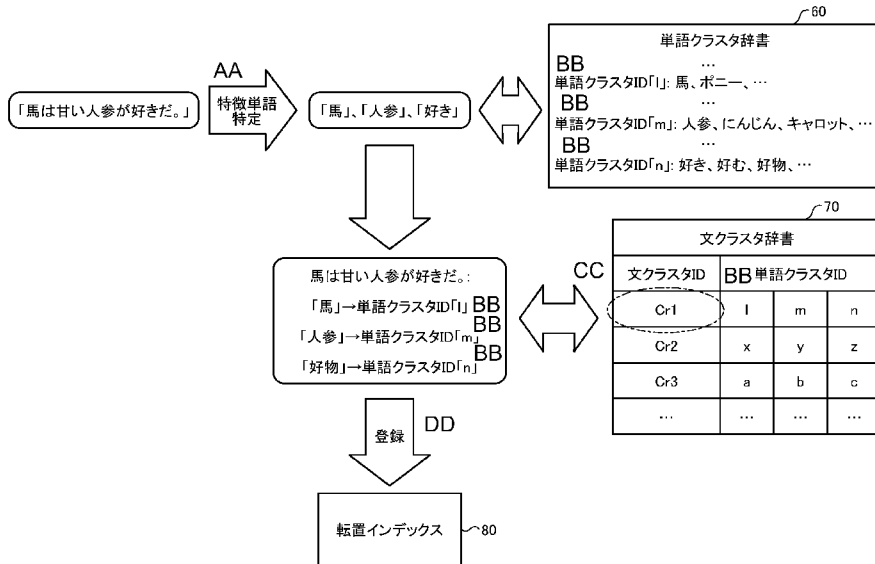
号 富士通株式会社内 Kanagawa (JP). ▲徳 ▼
田 隆志(TOKUDA, Takashi); 〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内 Kanagawa (JP).

(74) 代理人: 弁理士法人酒井国際特許事務所
(SAKAI INTERNATIONAL PATENT OFFICE);
〒1000013 東京都千代田区霞が関3丁目8番1号 虎の門三井ビルディング Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY,

(54) Title: INFORMATION PROCESSING PROGRAM, INFORMATION PROCESSING METHOD, AND INFORMATION PROCESSING DEVICE

(54) 発明の名称: 情報処理プログラム、情報処理方法および情報処理装置



60... Word cluster dictionary
70... Sentence cluster dictionary
80... Inverted indices
AA ... Feature word identification
BB ... Word cluster ID
CC ... Sentence cluster ID
DD ... Registration

(57) Abstract: Provided is an information processing device which acquires a plurality of sentences including a plurality of words. This information processing device applies processing to the plurality of sentences, the processing identifying a set of feature words from the plurality of words on the basis of sentence vectors of the sentences including the plurality of words and word vectors of the plurality of words. The information processing device classifies the plurality of sentences such that sentences including the same set of feature words are included in the same cluster.

(57) 要約: 情報処理装置は、複数の単語を有する複数の文を取得する。情報処理装置は、複数の単語を有する文の文ベクトルと、複数の単語の単語ベクトルとを基にして、複数の単語から特徴単語の組を特定する処理を、複数の文に対して実行する。情報処理装置は、同一の特徴単語の組を有する文が、同一のクラスターに含まれるように、前記複数の文を分類する。

MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK,
SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類：

一 国際調査報告 (条約第21条(3))

明 細 書

発明の名称：

情報処理プログラム、情報処理方法および情報処理装置

技術分野

[0001] 本発明は、情報処理プログラム等に関する。

背景技術

[0002] DB (Data Base) にはテキスト等の膨大なデータが登録されており、かかるDBから、利用者に指定された検索クエリに類似するデータを適切に検索することが求められている。以下では、テキストを、複数の単語を有する文として説明を行う。

[0003] 従来技術では、テキストがDBに登録された時に転置インデックスを設定し、検索クエリを受け付けた場合に、データ検索を実行する。たとえば、従来技術では、事前準備において、テキストがDBに登録された時、各文のベクトル（以下、文ベクトル）を算出して、類似した文ベクトルを同一のクラスタに分類する。従来技術では、同一のクラスタに含まれる複数の文の位置とその代表ベクトルを対応付けて、転置インデックスに設定することで、検索する処理の効率化をはかっている。

[0004] 従来技術では、文ベクトルを算出する場合に、文を構成する複数の単語の各単語ベクトルをそれぞれ算出し、全単語の単語ベクトルを積算することで、文ベクトルを算出している。

先行技術文献

特許文献

[0005] 特許文献1：国際公開第2020/095357号

特許文献2：特開2019-101993号公報

発明の概要

発明が解決しようとする課題

[0006] 文には、名詞や動詞、形容詞、助詞などの様々な単語が含まれており、従

来技術のように、単に、文に含まれる全単語の単語ベクトルを積算した文ベクトルを算出すると、文の特徴を明確に示す文ベクトルとはならない場合がある。このような文ベクトルを用いて、クラスタリングを実行すると、本来であれば、別々のクラスタに分類すべき複数の文が、同一のクラスタに分類されてしまう場合がある。

[0007] 1つの側面では、本発明は、テキストを適切にクラスタリングすることができる情報処理プログラム、情報処理方法および情報処理装置を提供することを目的とする。

課題を解決するための手段

[0008] 第1の案では、コンピュータに次の処理を実行させる。コンピュータは、複数の単語を有する複数の文を取得する。コンピュータは、複数の単語を有する文の文ベクトルと、複数の単語の単語ベクトルとを基にして、複数の単語から特徴単語の組を特定する処理を、複数の文に対して実行する。コンピュータは、同一の特徴単語の組を有する文が、同一のクラスタに含まれるように、前記複数の文を分類する。

発明の効果

[0009] テキストを適切にクラスタリングすることができる。

図面の簡単な説明

[0010] [図1]図1は、文の特徴単語を特定する処理を説明するための図である。

[図2]図2は、クラスタリングする処理を説明するための図である。

[図3]図3は、検索フェーズの処理を説明するための図(1)である。

[図4]図4は、検索フェーズの処理を説明するための図(2)である。

[図5]図5は、本実施例に係る情報処理装置の構成を示す機能ブロック図である。

[図6]図6は、単語ベクトル辞書のデータ構造の一例を示す図である。

[図7]図7は、転置インデックスのデータ構造を示す図である。

[図8]図8は、本実施例に係る情報処理装置の準備フェーズの処理手順を示すフローチャートである。

[図9]図9は、本実施例に係る情報処理装置の検索フェーズの処理手順を示すフローチャートである。

[図10]図10は、複数の文に基づく検索処理の処理手順を示すフローチャートである。

[図11]図11は、情報処理装置のその他の処理を説明するための図である。

[図12]図12は、実施例の情報処理装置と同様の機能を実現するコンピュータのハードウェア構成の一例を示す図である。

発明を実施するための形態

[0011] 以下に、本願の開示する情報処理プログラム、情報処理方法および情報処理装置の実施例を図面に基づいて詳細に説明する。なお、この実施例によりこの発明が限定されるものではない。

実施例

[0012] 本実施例に係る情報処理装置は、準備フェーズの処理と、検索フェーズの処理とを行う。まず、情報処理装置が実行する準備フェーズの処理について説明する。たとえば、準備フェーズには、文の特徴単語を特定する処理、文をクラスタリングする処理が含まれる。

[0013] 図1は、文の特徴単語を特定する処理を説明するための図である。図1では、テキストDB50に登録された、文「馬は甘い人参が好きだ。」から、特徴単語を特定する場合について説明する。

[0014] 情報処理装置は、単語と単語のベクトルとの関係を定義した単語ベクトル辞書を基にして、文「馬は甘い人参が好きだ。」に含まれる各単語のベクトルを特定する。以下の説明では、単語のベクトルを「単語ベクトル」と表記する。たとえば、「馬」の単語ベクトルを $wv-a$ とする。「甘い」の単語ベクトルを $wv-b$ とする。「人参」の単語ベクトルを $wv-c$ とする。「好き」の単語ベクトルを $wv-d$ とする。単語「は」、「が」、「だ」の単語ベクトルの図示を省略する。情報処理装置は、文「馬は甘い人参が好きだ。」の各単語の単語ベクトルを積算することで、文「馬は甘い人参が好きだ。」の文ベクトル $sv-1$ を算出する。

[0015] 情報処理装置は、文ベクトル $s_v 1$ と、単語ベクトル $w_v - a \sim w_v - d$ とのコサイン類似度をそれぞれ算出し、コサイン類似度を基にして、文ベクトル $s_v 1$ から乖離する単語ベクトルの単語を、「特徴単語」として特定する。たとえば、情報処理装置は、文ベクトル $s_v 1$ とのコサイン類似度が閾値以上となる単語ベクトルの単語を、特徴単語とする。

[0016] 図1では、文ベクトル $s_v 1$ 及び単語ベクトル $w_v - a$ のコサイン類似度、文ベクトル $s_v 1$ 及び単語ベクトル $w_v 1 - c$ のコサイン類似度、文ベクトル $s_v 1$ 及び単語ベクトル $w_v - d$ のコサイン類似度をそれぞれ、閾値以上とする。そうすると、情報処理装置は、単語ベクトル $w_v - a$ の単語「馬」、単語ベクトル $w_v 1 - c$ の単語「人参」、単語ベクトル $w_v - d$ の単語「好き」を、特徴単語として特定する。

[0017] 続いて、情報処理装置は、図1の処理で特定した特徴単語を基にして、文をクラスタリングする。図2は、クラスタリングする処理を説明するための図である。情報処理装置は、単語クラスタ辞書60を基にして、各特徴単語の単語クラスタIDを特定する。単語クラスタ辞書60は、複数の単語が、複数のクラスタに分類されており、各クラスタには、それぞれ単語クラスタIDが設定されている。情報処理装置は、特徴単語の属するクラスタに設定された単語クラスタIDを特定し、特定した単語クラスタIDを、特徴単語の単語クラスタIDとする。

[0018] たとえば、情報処理装置は、単語クラスタ辞書60を基にして、特徴単語「馬」を含むクラスタに設定された単語クラスタID「l」を特定し、特徴単語「馬」の単語クラスタIDとする。情報処理装置は、単語クラスタ辞書60を基にして、特徴単語「人参」を含むクラスタに設定された単語クラスタID「m」を特定し、特徴単語「人参」の単語クラスタIDとする。情報処理装置は、単語クラスタ辞書60を基にして、特徴単語「好き」を含むクラスタに設定された単語クラスタID「n」を特定し、特徴単語「好き」の単語IDとする。

[0019] 情報処理装置が上記処理を実行することで、特徴単語「馬」、「人参」、

「好き」に対応する単語クラスタID「l」、「m」、「n」を特定する。情報処理装置は、かかる単語クラスタIDの組「l」、「m」、「l」を、文「馬は甘い人参が好きだ。」に対応する単語クラスタIDの組として設定する。

[0020] 続いて、情報処理装置は、文に設定した単語クラスタIDの組と、文クラスタ辞書70とを基にして、文の属する文クラスタを特定する。ここで、文クラスタ辞書70は、文のクラスタを識別する文クラスタIDと、単語クラスタIDの組とを対応付ける。たとえば、単語クラスタID「l」、「m」、「n」に対応する文クラスタIDは、「Cr1」である。このため、情報処理装置は、文「馬は甘い人参が好きだ。」が属する文クラスタの文クラスタIDを「Cr1」と特定する。

[0021] 情報処理装置は、特定した文「馬は甘い人参が好きだ。」の文クラスタIDと、テキストDB50の文「馬は甘い人参が好きだ。」の位置とを対応付けて、転置インデックス80に登録する。

[0022] 情報処理装置は、テキストDB50に登録された各文に対して、上記処理を繰り返し実行し、各文の文クラスタIDと、各文の位置との関係を、転置インデックス80に登録する。

[0023] 上記のように、本実施例に係る情報処理装置は、文に含まれる複数の単語から特徴単語の組を特定し、特徴単語の組と、文クラスタ辞書70とを基にして、文の属するクラスタの文クラスタIDを特定する。これによって、文を適切にクラスタリングすることができる。

[0024] 次に、情報処理装置が実行する検索フェーズの処理について説明する。図3は、検索フェーズの処理を説明するための図(1)である。図3の説明では、検索クエリq1として、単一な文「馬は甘い人参が好物だ。」が指定された場合について説明する。

[0025] 情報処理装置は、検索クエリq1を受け付けると、検索クエリq1の文「馬は甘い人参が好物だ。」から、特徴単語「馬」、「人参」、「好物」を特定する。情報処理装置が、文に含まれる複数の単語から、特徴単語を特定す

る処理は、図 1 で説明した処理と同様である。

- [0026] 情報処理装置は、各特徴単語と、単語クラスタ辞書 60 とを基にして、各特徴単語の単語クラスタ ID を特定する。たとえば、情報処理装置は、単語クラスタ辞書 60 を基にして、特徴単語「馬」を含むクラスタに設定された単語クラスタ ID 「l」を特定し、特徴単語「馬」の単語クラスタ ID とする。情報処理装置は、単語クラスタ辞書 60 を基にして、特徴単語「人参」を含むクラスタに設定された単語クラスタ ID 「m」を特定し、特徴単語「人参」の単語クラスタ ID とする。情報処理装置は、単語クラスタ辞書 60 を基にして、特徴単語「好物」を含むクラスタに設定された単語クラスタ ID 「n」を特定し、特徴単語「好物」の単語 ID とする。
- [0027] 情報処理装置が上記処理を実行することで、特徴単語「馬」、「人参」、「好物」に対応する単語クラスタ ID 「l」、「m」、「n」を特定する。情報処理装置は、かかる単語クラスタ ID の組「l」、「m」、「n」を、検索クエリ q 1 の文「馬は甘い人参が好物だ。」に対応する単語クラスタ ID の組として設定する。
- [0028] 情報処理装置は、検索クエリ q 1 の文に設定した単語クラスタ ID の組と、文クラスタ辞書 70 とを基にして、検索クエリ q 1 の文の属する文クラスタを特定する。たとえば、単語クラスタ ID 「l」、「m」、「n」に対応する文クラスタ ID は、「C r 1」である。このため、情報処理装置は、検索クエリ q 1 の文「馬は甘い人参が好物だ。」が属する文クラスタの文クラスタ ID を「C r 1」と特定する。
- [0029] 情報処理装置は、検索クエリ q 1 の文に対応する文クラスタ ID（たとえば、「C r 1」）と、転置インデックス 80 とを基にして、文クラスタ ID に属するテキスト DB 50 上の文の位置を特定する。情報処理装置は、特定した位置から、文を抽出し、抽出した文を、検索結果として出力する。
- [0030] 上記のように、情報処理装置は、検索クエリ q 1 に含まれる複数の単語から特徴単語の組を特定し、特徴単語の組と、文クラスタ辞書 70 とを基にして、検索クエリ q 1 に対応する文クラスタ ID を特定する。そして、情報処

理装置は、事前に作成しておいた転置インデックス80と、検索クエリ q_1 に対応する文クラスターIDとを基にして、検索を行う。これによって、検索クエリ q_1 に対応する文を適切に検索することができる。

[0031] 続いて、図4の説明を行う。図4は、検索フェーズの処理を説明するための図(2)である。図4の説明では、検索クエリとして、段落や項目などの複数の文が指定された場合について説明する。たとえば、検索クエリ q_2 には、複数の文「このプログラムの特徴は以下の内容です。・・・複数のサブプロで構成されている。・・・その機能が特徴です。・・・高速化の効果を実現した。・・・」が含まれている。

[0032] 情報処理装置は、検索クエリ q_2 に含まれる複数の文の文ベクトルをそれぞれ算出する。たとえば、情報処理装置は、1つの文に含まれる複数の単語の単語ベクトルを積算することで、1つの文の文ベクトルを算出する。または、情報処理装置は、図1、2で説明した処理を実行して、文の各特徴単語の単語クラスターIDの組を特定し、かかる単語クラスターIDの組を、文ベクトルとして利用してもよい。

[0033] 図4に示す例では、文「その機能が特徴です」の文ベクトルを文ベクトル sv_{2-1} とする。文「このプログラムの特徴は以下の内容です。」の文ベクトルを文ベクトル sv_{2-2} とする。文「複数のサブプロで構成されている。」の文ベクトルを文ベクトル sv_{2-3} とする。文「高速化の効果を實現した。」の文ベクトルを文ベクトル sv_{2-4} とする。

[0034] 情報処理装置は、検索クエリ q_2 に含まれる複数の文の文ベクトルを積算することで、文章ベクトル dv_1 を算出する。

[0035] 情報処理装置は、文章ベクトル dv_1 と、文ベクトル $sv_{2-1} \sim sv_{2-4}$ とのコサイン類似度をそれぞれ算出し、コサイン類似度を基にして、文章ベクトル dv_1 から乖離する文ベクトルの文を、「特徴文」として特定する。たとえば、情報処理装置は、文章ベクトル dv_1 とのコサイン類似度が閾値以上となる文ベクトルの文を、特徴文とする。

[0036] 図4では、文章ベクトル dv_1 及び文ベクトル sv_{1-1} のコサイン類似

度、文章ベクトル $d_v 1$ 及び文ベクトル $s_v 1-3$ のコサイン類似度、文章ベクトル $d_v 1$ 及び文ベクトル $s_v 1-4$ のコサイン類似度をそれぞれ、閾値以上とする。そうすると、情報処理装置は、文ベクトル $s_v 1-1$ の文、文ベクトル $s_v 1-3$ の文、文ベクトル $s_v 1-4$ の単語を、特徴文として特定する。

[0037] 情報処理装置は、特徴文それぞれについて、図3で説明した処理を実行することで、テキストDB50から、特徴文に対応する文を検索する。図4に示す例では、特徴文「その機能が特徴です。」に対応する検索候補として、文X1、文X2、文X3、文X4が検索されている。特徴文「複数のサブプロで構成されている。」に対応する検索候補として、文X2、文X3、文X6、文X10が検索されている。特徴文「高速化を実現した。」に対応する検索候補として、文X2、文X3、文X7、文X22が検索されている。

[0038] 情報処理装置は、各特徴文の検索候補で共通する文を、最終的な検索結果として特定する。図4に示す例では、各検索候補に共通する文X2、文X3を最終的な検索結果として出力する。

[0039] 上記のように、情報処理装置は、検索クエリ $q 2$ に複数の文が含まれる場合には、特徴文を特定し、特徴文に対応する検索結果で共通する文を最終的な検索結果として特定する。これによって、検索クエリ $q 2$ に複数の文が含まれていても、検索クエリ $q 2$ に対応する文を効率的に検索することができる。

[0040] 次に、上述した処理を実行する情報処理装置の構成例について説明する。図5は、本実施例に係る情報処理装置の構成を示す機能ブロック図である。図5に示すように、情報処理装置100は、通信部110、入力部120、表示部130、記憶部140、制御部150を有する。

[0041] 通信部110は、有線又は無線で外部装置等に接続され、外部装置等との間で情報の送受信を行う。たとえば、通信部110は、NIC (Network Interface Card) 等によって実現される。通信部110は、図示しないネットワークに接続されていてもよい。

- [0042] 入力部120は、各種の情報を、情報処理装置100に入力する入力装置である。入力部120は、キーボードやマウス、タッチパネル等に対応する。たとえば、ユーザは、入力部120を操作して、文、検索クエリ等のデータ等を入力してもよい。
- [0043] 表示部130は、制御部150から出力される情報を表示する表示装置である。表示部130は、液晶ディスプレイ、有機EL (Electro Luminescence) ディスプレイ、タッチパネル等に対応する。たとえば、検索クエリの検索結果が、表示部130に表示される。
- [0044] 記憶部140は、単語ベクトル辞書40、テキストDB50と、単語クラスタ辞書60、文クラスタ辞書70、転置インデックス80を有する。記憶部140は、たとえば、RAM (Random Access Memory)、フラッシュメモリ (Flash Memory) 等の半導体メモリ素子、または、ハードディスク、光ディスク等の記憶装置によって実現される。
- [0045] 単語ベクトル辞書40は、単語に割り当てられた符号、単語ベクトルを定義するテーブルである。図6は、単語ベクトル辞書のデータ構造の一例を示す図である。図6に示すように、この単語ベクトル辞書40は、符号、単語、単語ベクトル(1)～(7)を有する。符号は、単語に割り当てられる符号(Code)である。単語は、文字列に含まれる単語である。単語ベクトル(1)～(7)は、単語に割り当てられたベクトルである。単語ベクトルの第 n 成分を単語ベクトル(n)と表記する($n=1\sim7$)。
- [0046] テキストDB50は、複数の文を格納するデータベースである。たとえば、テキストDB50は、複数のレコードが含まれる。一つのレコードには、複数の文が含まれる。
- [0047] 単語クラスタ辞書60は、複数の単語が、複数のクラスタに分類されており、各クラスタには、それぞれ単語クラスタIDが設定されている。同一のクラスタに分類される複数の単語は、各単語の単語ベクトルのコサイン類似度が閾値以上となる。単語クラスタ辞書60に関するその他の説明は、図2で行った説明と同様である。

- [0048] 文クラスタ辞書 70 は、文のクラスタを識別する文クラスタ ID と、単語クラスタ ID の組とを対応付ける。同一の文クラスタに属する複数の文には、同一の文クラスタ ID が設定される。文クラスタ辞書 70 に関するその他の説明は、図 2 で行った説明と同様である。
- [0049] 転置インデックス 80 は、文クラスタ ID と、文クラスタ ID に属する文の位置（テキスト DB 50 上の位置）とを対応付ける。図 7 は、転置インデックスのデータ構造を示す図である。図 7 に示すように、この転置インデックス 80 は、文クラスタ ID と対応付けて、レコードポインタと位置ポインタとの組が複数設定される。レコードポインタは、該当レコードの位置を示す。レコードポインタによって示される位置には、テキスト DB 50 の先頭単語からレコードの先頭単語までの単語の数（オフセット）によって定義される。位置ポインタは、該当文の位置を示す。位置ポインタは、該当する文を含むレコードの先頭の単語から、該当文の先頭単語までのオフセットによって定義される。
- [0050] たとえば、文「馬は甘い人参が好きだ。」の文クラスタ ID が「C r 1」であり、文「馬は甘い人参が好きだ。」がレコード R 1 に含まれる場合には、次のように設定される。すなわち、文クラスタ ID 「C r 1」に対応するレコードポインタ（1）には、レコード R 1 のオフセットが設定される。文クラスタ ID 「C r 1」に対応する位置ポインタ（1）には、文「馬は甘い人参が好きだ。」のオフセットが設定される。
- [0051] なお、転置インデックス 80 のデータ構造は、図 7 に限定されるものではなく、シンプルに、文クラスタ ID と、該当文クラスタ ID に属する各文のオフセットとを対応付けるものであってもよい。
- [0052] 図 5 の説明に戻る。制御部 150 は、取得部 151 と、前処理部 152 と、検索部 153 とを有する。制御部 150 は、たとえば、CPU (Central Processing Unit) や MPU (Micro Processing Unit) により実現される。また、制御部 150 は、例えば ASIC (Application Specific Integrated Circuit) や FPGA (Field Programmable Gate Array) 等の集積回

路により実行されてもよい。

- [0053] 取得部 151 は、通信部 110 または入力部 120 を介して、各種の情報を取得する。たとえば、取得部 151 は、レコードの情報を取得した場合には、取得したレコードの情報を、テキスト DB 50 に登録する。
- [0054] 前処理部 152 は、上述した準備フェーズの処理を実行する。前処理部 152 は、テキスト DB から文を取得し、図 1 で説明した処理を実行することで、文に含まれる特徴単語を特定する。前処理部 152 は、特徴単語を特定すると、図 2 で説明したように文のクラスタリングを実行する。たとえば、前処理部 152 は、単語クラスタ辞書 60 を基にして、各特徴単語の属するクラスタに設定された単語クラスタ ID を特定する。前処理部 152 は、特定した単語クラスタ ID の組と、文クラスタ辞書 70 とを基にして、文が属する文クラスタ ID を特定する。前処理部 152 は、文のクラスタ ID と、文の位置を特定可能なレコードポイントおよび位置ポイントとを対応付けて、転置インデックス 80 に設定する。
- [0055] 前処理部 152 は、テキスト DB 50 に登録された各文について、上記処理を繰り返し実行する。
- [0056] 検索部 153 は、上述した検索フェーズの処理を実行する。検索部 153 は、通信部 110 または入力部 120 を介して、検索クエリを取得する。検索部 153 は、検索クエリに 1 つの文が含まれているか、複数の文が含まれているかを判定する。
- [0057] 検索クエリに単一の文が含まれている場合について説明する。検索部 153 は、検索クエリに 1 つの文が含まれている場合には、図 3 で説明した処理を実行する。たとえば、検索部 153 は、検索クエリに含まれる文から、特徴単語を特定する。検索部 153 は、各特徴単語と、単語クラスタ辞書 60 とを基にして、各特徴単語の単語クラスタ ID の組を特定する。検索部 153 は、単語クラスタ ID の組と、文クラスタ辞書 70 とを基にして、検索クエリに対応する文クラスタ ID を特定する。
- [0058] 検索部 153 は、検索クエリに対応する文クラスタ ID と、転置インデッ

クス80とを基にして、文クラスターIDに対応するレコードポインタおよび位置ポインタの組を特定する。検索部153は、特定したレコードポインタおよび位置ポインタの組に対応する文（複数の文）を、テキストDB50から取得し、検索結果として、表示部130に表示させる。検索部153は、検索結果を、外部装置に通知してもよい。

[0059] 続いて、検索クエリに段落や項目などの複数の文が含まれている場合について説明する。検索部153は、検索クエリに複数の文が含まれている場合には、図4で説明した処理を実行する。検索部153は、検索クエリに含まれる複数の文から特徴文を特定する。検索部153は、各特徴文に対応する単語クラスターIDの組と、文クラスター辞書70とを基にして、各特徴文の文クラスターIDを特定する。検索部153は、各特徴文の文クラスターIDと、転置インデックス80とを基にして、各特徴文に対応する複数の文（検索結果）をテキストDB50から取得する。検索部153は、各特徴文に対応する検索結果に共通する文を、最終的な検索結果として検索する。検索部153は、検索結果を、表示部130に表示させる。検索部153は、検索結果を、外部装置に通知してもよい。

[0060] 次に、本実施例に係る情報処理装置の処理手順の一例について説明する。図8は、本実施例に係る情報処理装置の準備フェーズの処理手順を示すフローチャートである。図8に示すように、情報処理装置100の前処理部152は、テキストDB50から未処理の文を取得する（ステップS101）。前処理部152は、単語ベクトル辞書40を基にして、文に含まれる複数の単語の単語ベクトルを積算し、文ベクトルを算出する（ステップS102）。

[0061] 前処理部152は、文ベクトルと、各単語の単語ベクトルとのコサイン類似度を基にして、特徴単語を特定する（ステップS103）。前処理部152は、単語クラスター辞書60を基にして、特徴単語に対する単語クラスターIDを特定する（ステップS104）。

[0062] 前処理部152は、文の特徴単語に対する単語クラスターIDの組と、文ク

ラスト辞書70とを基にして、文の属するクラスタの文クラスタIDを特定する（ステップS105）。文の位置情報（レコードポインタおよび位置ポインタの組）と、文クラスタIDとを対応付けて、転置インデックス80に登録する（ステップS106）。

[0063] 前処理部152は、未処理の文がテキストDB50に存在する場合には（ステップS107, Yes）、ステップS101に移行する。一方、前処理部152は、未処理の文がテキストDB50に存在しない場合には（ステップS107, No）、準備フェーズの処理を終了する。

[0064] 図9は、本実施例に係る情報処理装置の検索フェーズの処理手順を示すフローチャートである。図9に示すように、情報処理装置100の検索部153は、検索クエリを受け付ける（ステップS201）。検索部153は、検索クエリに複数の文が含まれるか否かを判定する（ステップS202）。

[0065] 検索部153は、検索クエリに複数の文が含まれない場合には（ステップS202, No）、ステップS203に移行する。検索部153は、単語ベクトル辞書40を基にして、検索クエリの文に含まれる複数の単語の単語ベクトルを積算し、文ベクトルを算出する（ステップS203）。

[0066] 検索部153は、文ベクトルと、各単語の単語ベクトルとのコサイン類似度を基にして、検索クエリの文に含まれる特徴単語を特定する（ステップS204）。検索部153は、単語クラスタ辞書60を基にして、特徴単語に対する単語クラスタIDを特定する（ステップS205）。

[0067] 検索部153は、検索クエリの文の特徴単語に対する単語クラスタIDの組と、文クラスタ辞書70とを基にして、検索クエリの文の属するクラスタの文クラスタIDを特定する（ステップS206）。検索部153は、検索クエリの文の文クラスタIDと、転置インデックス80とを基にして、文クラスタIDに対応する文の位置情報を特定する（ステップS207）。

[0068] 検索部153は、位置情報に対応する位置の文を、テキストDB50から取得する（ステップS208）。検索部153は、検索結果を出力する（ステップS209）。

- [0069] 一方、検索部153は、ステップS202において、検索クエリに複数の文が含まれる場合には（ステップS202、Yes）、ステップS210に移行する。検索部153は、複数の文に基づく検索処理を実行し（ステップS210）、ステップS209に移行する。
- [0070] ここで、図9のステップS210に示した複数の文に基づく検索処理の処理手順の一例について説明する。図10は、複数の文に基づく検索処理の処理手順を示すフローチャートである。図10に示すように、情報処理装置100の検索部153は、検索クエリに含まれる複数の文から未選択の文を選択する（ステップS301）。
- [0071] 検索部153は、単語ベクトル辞書40を基にして、選択した文に含まれる複数の単語の単語ベクトルを積算し、文ベクトルを算出する（ステップS302）。検索部153は、文ベクトルと、各単語の単語ベクトルとのコサイン類似度を基にして、文に含まれる特徴単語を特定する（ステップS303）。検索部153は、単語クラスタ辞書60を基にして、特徴単語に対する単語クラスタIDを特定する（ステップS304）。
- [0072] 検索部153は、文の特徴単語に対する単語クラスタIDの組と、文クラスタ辞書70とを基にして、文の属するクラスタの文クラスタIDを特定する（ステップS305）。検索部153は、文の文クラスタIDと、転置インデックス80とを基にして、文クラスタIDに対応する文の位置情報を特定する（ステップS306）。
- [0073] 検索部153は、位置情報に対応する位置の文（検索結果）を、テキストDB50から取得する（ステップS307）。
- [0074] 検索部153は、検索クエリに未処理の文が存在する場合には（ステップS308、Yes）、ステップS301に移行する。検索部153は、検索クエリに未処理の文が存在しない場合には（ステップS308、No）、検索クエリに含まれる各文の検索結果に共通する文を最終的な検索結果に設定し（ステップS309）、複数の文に基づく検索処理を終了する。
- [0075] 次に、本実施例に係る情報処理装置100の効果について説明する。情報

処理装置１００は、文に含まれる複数の単語から特徴単語の組を特定し、特徴単語の組と、文クラスタ辞書７０とを基にして、文の属するクラスタの文クラスタＩＤを特定する。これによって、文を適切にクラスタリングすることができる。

[0076] 情報処理装置１００は、文の文ベクトルと、複数の単語の単語ベクトルとのコサイン類似度を算出し、文ベクトルとのコサイン類似度が閾値以上となる単語ベクトルの単語を、特徴単語として特定する。これによって、文ベクトルと乖離する特徴単語の特定することができる。

[0077] 情報処理装置１００は、文の属するクラスタの文クラスタＩＤと、文の位置情報とを対応付けることで、転置インデックス８０を生成する。かかる転置インデックス８０を用いることで、同一の文クラスタＩＤに属する複数の文の位置情報を容易に特定することが可能になる。

[0078] 情報処理装置１００は、一つの文を有する検索クエリ q_1 に含まれる複数の単語から特徴単語の組を特定し、特徴単語の組と、文クラスタ辞書７０とを基にして、検索クエリ q_1 に対応する文クラスタＩＤを特定する。そして、情報処理装置１００は、事前に作成しておいた転置インデックス８０と、検索クエリ q_1 に対応する文クラスタＩＤとを基にして、検索を行う。これによって、検索クエリ q_1 に対応する文を適切に検索することができる。

[0079] 情報処理装置１００は、検索クエリ q_2 に複数の文が含まれる場合には、特徴文を特定し、特徴文に対応する検索結果で共通する文を最終的な検索結果として特定する。これによって、検索クエリ q_2 に複数の文が含まれていても、検索クエリ q_2 に対応する文を効率的に検索することができる。

[0080] ところで、上述した情報処理装置１００の処理は一例であり、情報処理装置１００はその他の処理を実行してもよい。以下において、情報処理装置１００のその他の処理について説明する。

[0081] 情報処理装置１００の検索部１５３は、複数の文を有する検索クエリを受け付けた場合に、コサイン類似度が閾値以上となる複数の特徴文を特定し、各特徴文を用いた検索結果に共通する文を、最終的な検索結果として検出し

ていた。ここで、検索部153は、コサイン類似度と比較する閾値の変更を受け付けることで、特徴文の数を増加または減少させる処理を更に実行してもよい。

[0082] たとえば、検索部153は、入力部120から、特徴文を特定する際に利用する閾値の変更を受け付け、変更後の閾値の値と、特徴文との関係を、表示部130に表示させる処理を繰り返し実行する。閾値の値が大きくなるほど、特徴文の数は減少し、閾値の値が小さくなるほど、特徴文の数は増加する。検索部153は、入力部120から、確定指示を受け付けた場合に、特徴文を確定する。検索部153が、特徴文を確定した後の処理は、上述した従来技術と同様である。

[0083] さらに、上記の処理を情報処理装置100の検索部153の実行において、段落や項目などの検索クエリに含まれる特徴文の数を増減することで、検索候補の数を増減させるズームイン・アウト機能を実現させることができる。

[0084] また、上述した情報処理装置100では、テキストの文字列に関し、複数の文をクラスタリングする場合について説明したが、文の代わりに、ゲノムの塩基配列のタンパク質一次構造や、有機化合物の化学構造式の官能基一次構造等の情報についても同様に処理を実行することができる。たとえば、タンパク質の一次構造には、繰返し出現する複数の連続塩基酸配列Kmerが含まれている。以下の説明では、連続塩基酸配列Kmerをタンパク質の「基本構造」と表記する。なお、タンパク質の「基本構造」は、連続アミノ酸配列oligopeptideなどにより、表記される場合もある。

[0085] 図11は、情報処理装置のその他の処理を説明するための図である。図11に示すように、タンパク質の一次構造Pro1には、複数の基本構造「 α -Kmer」、「 β -Kmer」、「 γ -Kmer」、「 δ -Kmer」が含まれる。

[0086] 情報処理装置100は、基本構造と、基本構造のベクトルとを定義した基本構造ベクトル辞書を基にして、タンパク質の一次構造Pro1に含まれる

各基本構造のベクトルを特定する。たとえば、基本構造「 α -Kmer」のベクトルを v_1 とする。基本構造「 β -Kmer」のベクトルを v_2 とする。基本構造「 γ -Kmer」のベクトルを v_3 とする。基本構造「 δ -Kmer」のベクトルを v_4 とする。情報処理装置100は、タンパク質の一次構造Pro1に含まれる各基本構造のベクトルを積算することで、一次構造Pro1のベクトル t_{v_1} を算出する。

[0087] 情報処理装置100は、ベクトル t_{v_1} と、ベクトル $v_1 \sim v_4$ とのコサイン類似度をそれぞれ算出し、コサイン類似度を基にして、ベクトル t_{v_1} から乖離するベクトルの基本構造を、「特徴基本構造」として特定する。たとえば、情報処理装置は、ベクトル t_{v_1} とのコサイン類似度が閾値以上となるベクトルの基本構造を、特徴基本構造とする。

[0088] 図11では、ベクトル t_{v_1} 及びベクトル v_1 のコサイン類似度、ベクトル t_{v_1} 及びベクトル v_3 のコサイン類似度、ベクトル t_{v_1} 及びベクトル v_4 のコサイン類似度をそれぞれ、閾値以上とする。そうすると、情報処理装置は、ベクトル v_1 の基本構造「 α -Kmer」、ベクトル v_3 の基本構造「 γ -Kmer」、ベクトル v_4 の基本構造「 δ -Kmer」を、特徴基本構造として特定する。

[0089] 情報処理装置100は、上記の処理で特定した特徴基本構造を基にして、一次構造をクラスタリングする。具体的な処理は、図2と同様にして、各特徴基本構造に対して、基本構造クラスタIDを割り当て、基本構造クラスタIDの組に基づいて、一次構造のクラスタIDを特定する。その他の処理は、図2で説明した処理の特徴単語を、特徴基本構造に置き換え、文を、一次構造に置き換えた処理と同様である。同様に、検索する処理においても、図3、4で説明した処理の特徴単語を、特徴基本構造に置き換え、文を、一次構造に置き換えた処理と同様である。タンパク質一次構造のクラスタIDに対応付けた転置インデックスを用いて、複数の一次構造で構成された受容体の検索クエリに対し、類似した受容体を検索することができる。これを応用すると、バイオ医薬品のリガンドのターゲットである受容体に類似した受容

体を検索することができ、バイオ医薬の副反応を推定することができる。

[0090] 次に、上記実施例に示した情報処理装置１００と同様の機能を実現するコンピュータのハードウェア構成の一例について説明する。図１２は、実施例の情報処理装置と同様の機能を実現するコンピュータのハードウェア構成の一例を示す図である。

[0091] 図１２に示すように、コンピュータ２００は、各種演算処理を実行するＣＰＵ２０１と、ユーザからのデータの入力を受け付ける入力装置２０２と、ディスプレイ２０３とを有する。また、コンピュータ２００は、有線または無線ネットワークを介して、外部装置等との間でデータの授受を行う通信装置２０４と、インタフェース装置２０５とを有する。また、コンピュータ２００は、各種情報を一時記憶するＲＡＭ２０６と、ハードディスク装置２０７とを有する。そして、各装置２０１～２０７は、バス２０８に接続される。

[0092] ハードディスク装置２０７は、取得プログラム３０７ａ、前処理プログラム２０７ｂ、検索プログラム２０７ｃを有する。また、ＣＰＵ２０１は、各プログラム２０７ａ～２０７ｃを読み出してＲＡＭ２０６に展開する。

[0093] 取得プログラム２０７ａは、取得プロセス２０６ａとして機能する。前処理プログラム２０７ｂは、前処理プロセス２０６ｂとして機能する。検索プログラム２０７ｃは、検索プロセス２０６ｃとして機能する。

[0094] 取得プロセス２０６ａの処理は、取得部１５１の処理に対応する。前処理プロセス２０６ｂの処理は、前処理部１５２の処理に対応する。検索プロセス２０６ｃの処理は、検索部１５３の処理に対応する。

[0095] なお、各プログラム２０７ａ～２０７ｃについては、必ずしも最初からハードディスク装置２０７に記憶させておかなくても良い。例えば、コンピュータ２００に挿入されるフレキシブルディスク（ＦＤ）、ＣＤ－ＲＯＭ、ＤＶＤ、光磁気ディスク、ＩＣカードなどの「可搬用の物理媒体」に各プログラムを記憶させておく。そして、コンピュータ２００が各プログラム２０７ａ～２０７ｃを読み出して実行するようにしてもよい。

符号の説明

[0096]	4 0	単語ベクトル辞書
	5 0	テキストＤＢ
	6 0	単語クラスタ辞書
	7 0	文クラスタ辞書
	8 0	転置インデックス
	1 0 0	情報処理装置
	1 1 0	通信部
	1 2 0	入力部
	1 3 0	表示部
	1 4 0	記憶部
	1 5 0	制御部
	1 5 1	取得部
	1 5 2	前処理部
	1 5 3	検索部

請求の範囲

- [請求項1] 複数の単語を有する複数の文を取得し、
複数の単語を有する文の文ベクトルと、前記複数の単語の単語ベクトルとを基にして、前記複数の単語から特徴単語の組を特定する処理を、前記複数の文に対して実行し、
同一の特徴単語の組を有する文が、同一のクラスタに含まれるように、前記複数の文を分類する
処理をコンピュータに実行させることを特徴とする情報処理プログラム。
- [請求項2] 前記特徴単語の組を特定する処理は、前記文の文ベクトルと、前記複数の単語の単語ベクトルとのコサイン類似度を算出し、前記文ベクトルとのコサイン類似度が閾値以上となる単語ベクトルの単語を、特徴単語として特定することを特徴とする請求項1に記載の情報処理プログラム。
- [請求項3] 前記複数の文は記憶装置に登録され、
前記分類する処理の分類結果を基にして、前記文の属するクラスタを識別する識別情報と、前記記憶装置における前記文の位置情報とを対応付けた転置インデックス情報を生成する処理を更にコンピュータに実行させることを特徴とする請求項1に記載の情報処理プログラム。
- [請求項4] 複数の単語を有する検索文を受け付けた場合に、前記検索文の文ベクトルと、前記検索文の単語の単語ベクトルとを基にして、前記検索文の複数の単語から特徴単語の組を特定し、特定した特徴単語の組に対するクラスタの識別情報と、前記転置インデックスとを基にして、前記検索文に対応する文を前記記憶装置から検索する処理を更にコンピュータに実行させることを特徴とする請求項3に記載の情報処理プログラム。
- [請求項5] 複数の検索文を受け付けた場合に、前記複数の検索文の文ベクトル

を基にして、複数の特徴文を特定し、前記複数の特徴文から特徴単語の組をそれぞれ特定し、特徴文に対応する特徴単語の組を基にして、各特徴文のクラスタの識別情報を特定し、各特徴文のクラスタの識別情報と、前記転置インデックスとを基にして、各特徴文に対応する文であって、共通する文を前記記憶装置から検索する処理を更にコンピュータに実行させることを特徴とする請求項4に記載の情報処理プログラム。

[請求項6] 前記複数の検索文から特定される複数の特徴文の数を増加または減少させる処理を更にコンピュータに実行させることを特徴とする請求項5に記載の情報処理プログラム。

[請求項7] 複数の単語を有する複数の文を取得し、
複数の単語を有する文の文ベクトルと、前記複数の単語の単語ベクトルとを基にして、前記複数の単語から特徴単語の組を特定する処理を、前記複数の文に対して実行し、
同一の特徴単語の組を有する文が、同一のクラスタに含まれるように、前記複数の文を分類する
処理をコンピュータが実行することを特徴とする情報処理方法。

[請求項8] 前記特徴単語の組を特定する処理は、前記文の文ベクトルと、前記複数の単語の単語ベクトルとのコサイン類似度を算出し、前記文ベクトルとのコサイン類似度が閾値以上となる単語ベクトルの単語を、特徴単語として特定することを特徴とする請求項7に記載の情報処理方法。

[請求項9] 前記複数の文は記憶装置に登録され、
前記分類する処理の分類結果を基にして、前記文の属するクラスタを識別する識別情報と、前記記憶装置における前記文の位置情報とを対応付けた転置インデックス情報を生成する処理を更に実行することを特徴とする請求項7に記載の情報処理方法。

[請求項10] 複数の単語を有する検索文を受け付けた場合に、前記検索文の文ベ

クトルと、前記検索文の単語の単語ベクトルとを基にして、前記検索文の複数の単語から特徴単語の組を特定し、特定した特徴単語の組に対するクラスタの識別情報と、前記転置インデックスとを基にして、前記検索文に対応する文を前記記憶装置から検索する処理を更に実行することを特徴とする請求項 9 に記載の情報処理方法。

[請求項11] 複数の検索文を受け付けた場合に、前記複数の検索文の文ベクトルを基にして、複数の特徴文を特定し、前記複数の特徴文から特徴単語の組をそれぞれ特定し、特徴文に対応する特徴単語の組を基にして、各特徴文のクラスタの識別情報を特定し、各特徴文のクラスタの識別情報と、前記転置インデックスとを基にして、各特徴文に対応する文であって、共通する文を前記記憶装置から検索する処理を更に実行することを特徴とする請求項 10 に記載の情報処理方法。

[請求項12] 前記複数の検索文から特定される複数の特徴文の数を増加または減少させる処理を更に実行することを特徴とする請求項 11 に記載の情報処理方法。

[請求項13] 複数の単語を有する複数の文を取得し、
複数の単語を有する文の文ベクトルと、前記複数の単語の単語ベクトルとを基にして、前記複数の単語から特徴単語の組を特定する処理を、前記複数の文に対して実行し、
同一の特徴単語の組を有する文が、同一のクラスタに含まれるように、前記複数の文を分類する
処理を実行する制御部を有することを特徴とする情報処理装置。

[請求項14] 前記特徴単語の組を特定する処理は、前記文の文ベクトルと、前記複数の単語の単語ベクトルとのコサイン類似度を算出し、前記文ベクトルとのコサイン類似度が閾値以上となる単語ベクトルの単語を、特徴単語として特定することを特徴とする請求項 13 に記載の情報処理装置。

[請求項15] 前記複数の文は記憶装置に登録され、

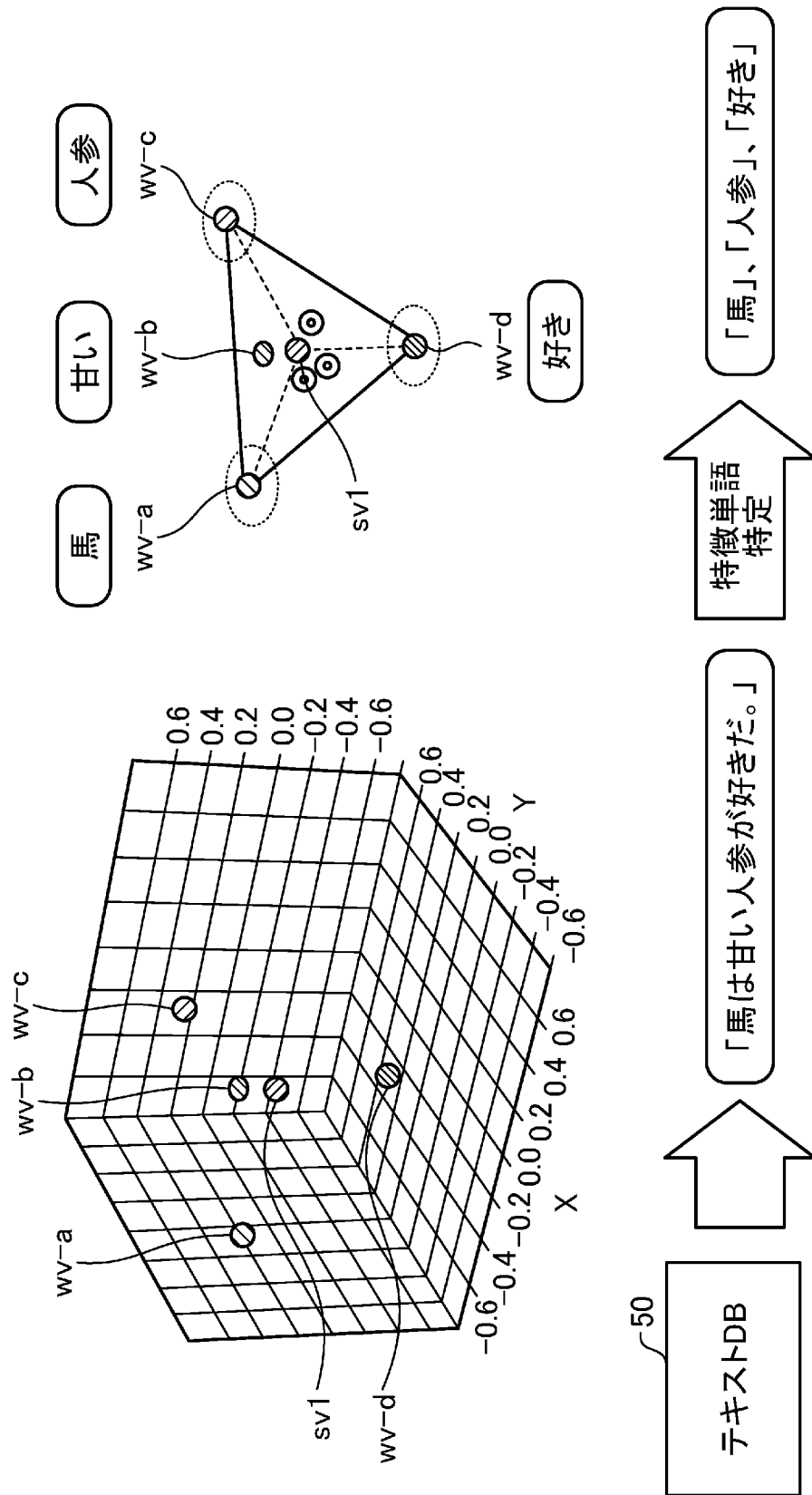
前記分類する処理の分類結果を基にして、前記文の属するクラスタを識別する識別情報と、前記記憶装置における前記文の位置情報とを対応付けた転置インデックス情報を生成する処理を更に実行することを特徴とする請求項 13 に記載の情報処理装置。

[請求項16] 複数の単語を有する検索文を受け付けた場合に、前記検索文の文ベクトルと、前記検索文の単語の単語ベクトルとを基にして、前記検索文の複数の単語から特徴単語の組を特定し、特定した特徴単語の組に対するクラスタの識別情報と、前記転置インデックスとを基にして、前記検索文に対応する文を前記記憶装置から検索する処理を更に実行することを特徴とする請求項 15 に記載の情報処理装置。

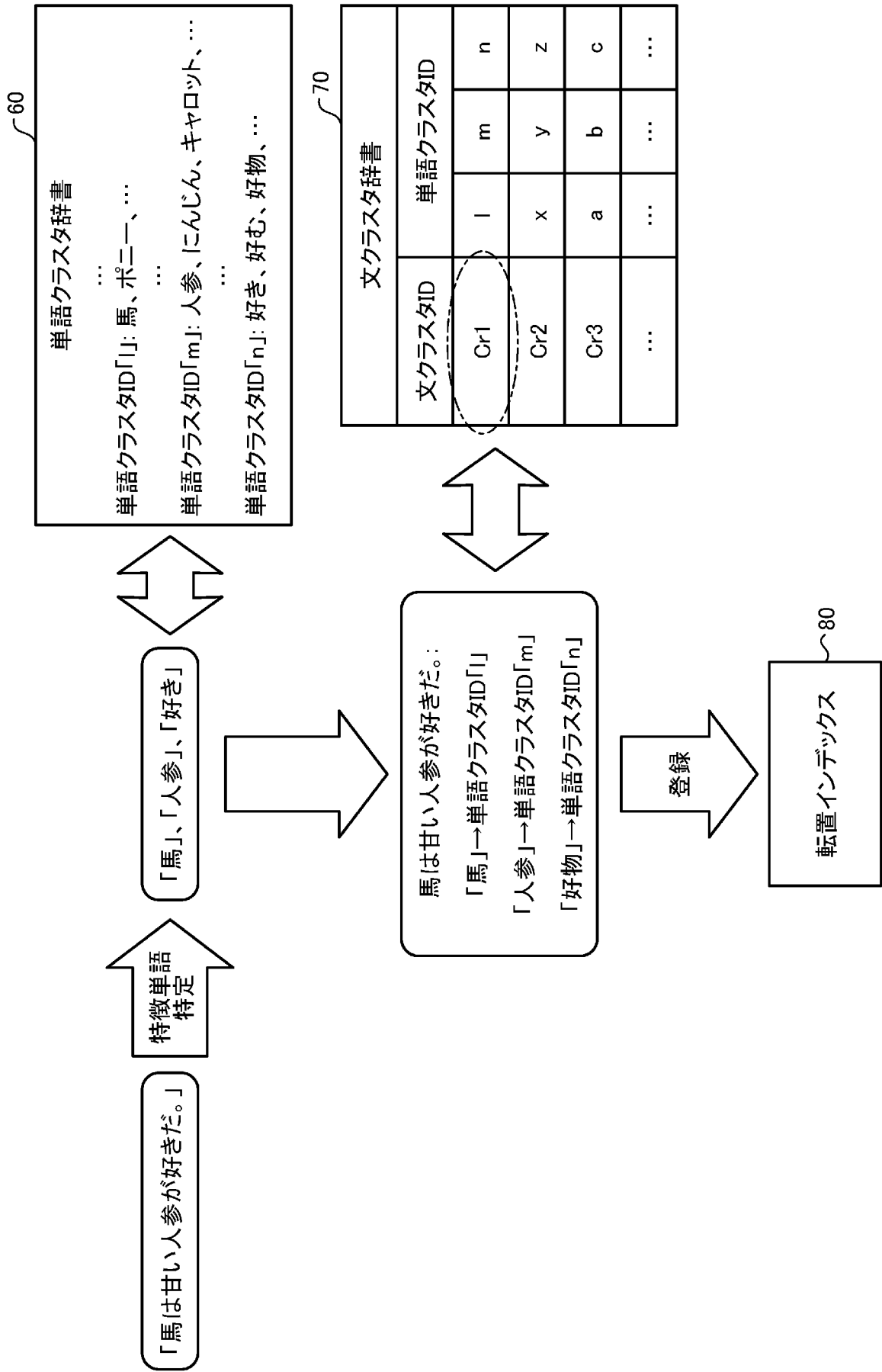
[請求項17] 複数の検索文を受け付けた場合に、前記複数の検索文の文ベクトルを基にして、複数の特徴文を特定し、前記複数の特徴文から特徴単語の組をそれぞれ特定し、特徴文に対応する特徴単語の組を基にして、各特徴文のクラスタの識別情報を特定し、各特徴文のクラスタの識別情報と、前記転置インデックスとを基にして、各特徴文に対応する文であって、共通する文を前記記憶装置から検索する処理を更に実行することを特徴とする請求項 16 に記載の情報処理装置。

[請求項18] 前記複数の検索文から特定される複数の特徴文の数を増加または減少させる処理を更に実行することを特徴とする請求項 17 に記載の情報処理装置。

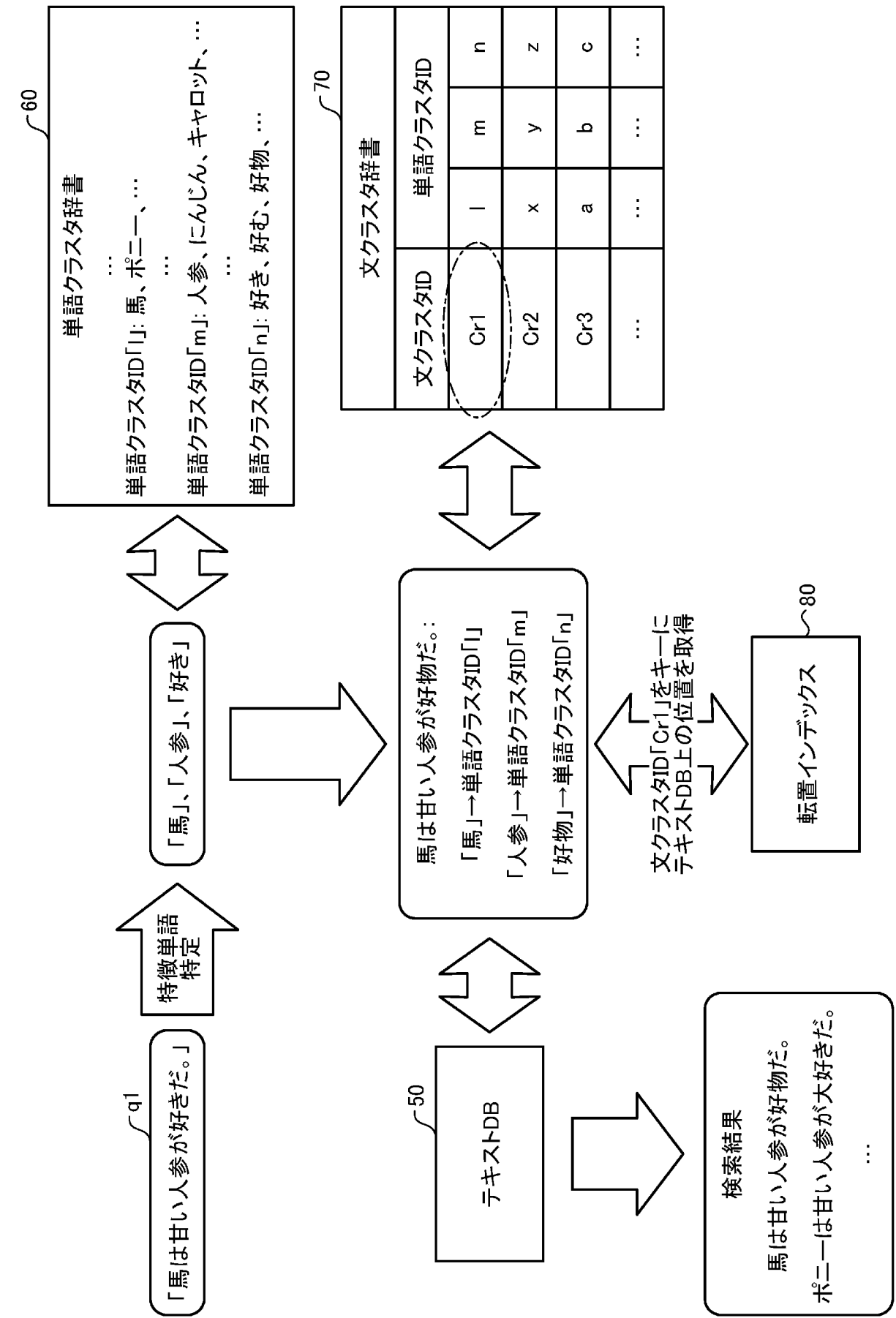
[図1]



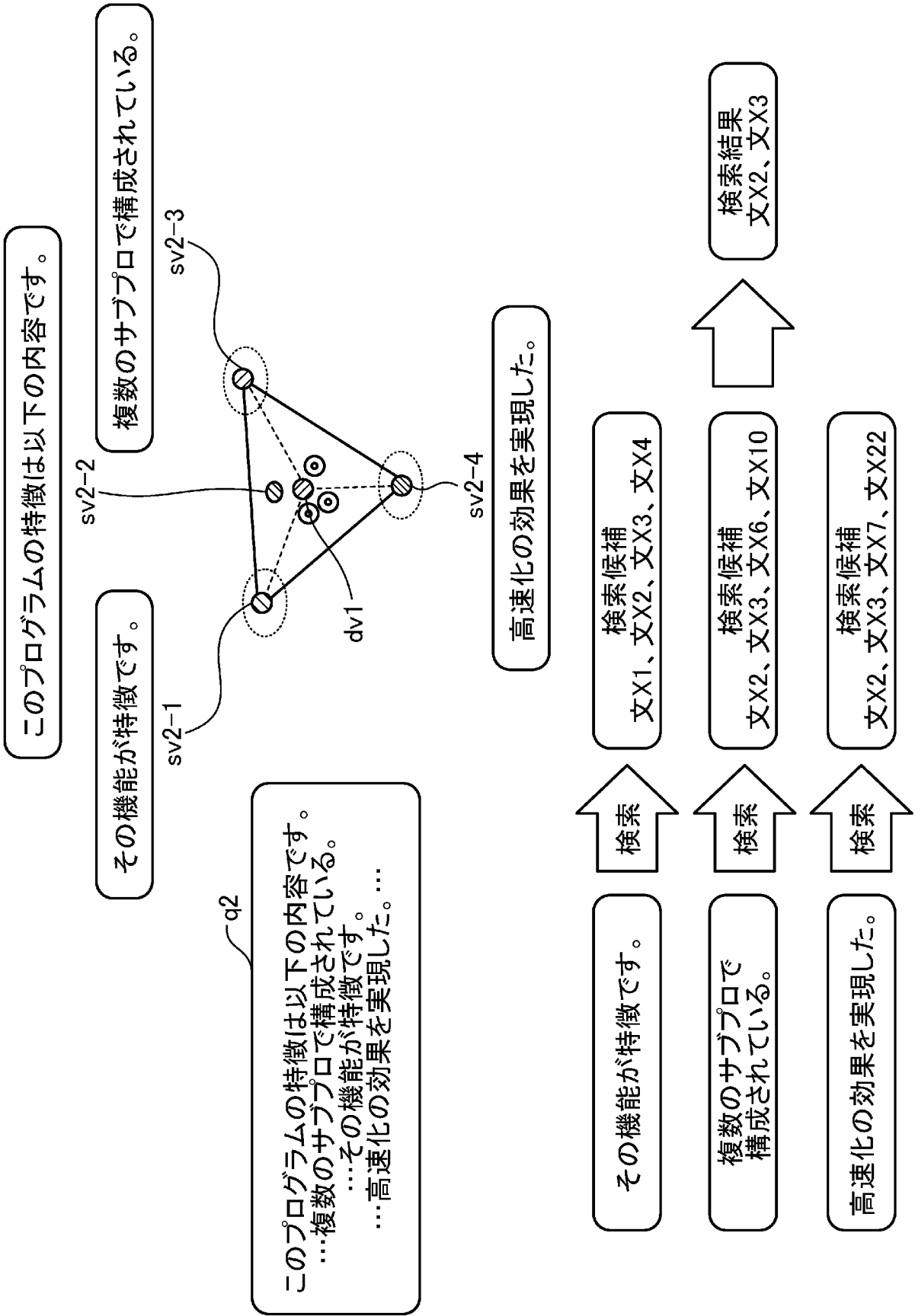
[図2]



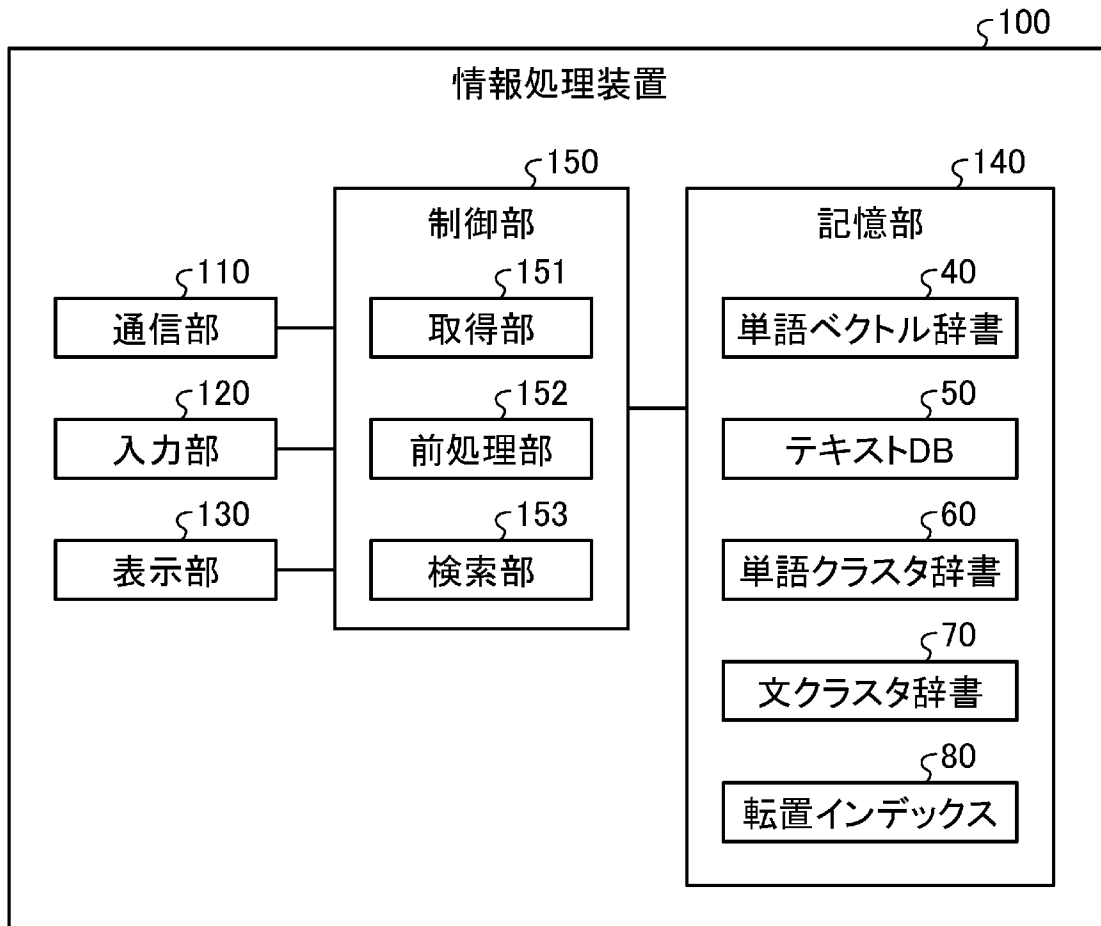
[図3]



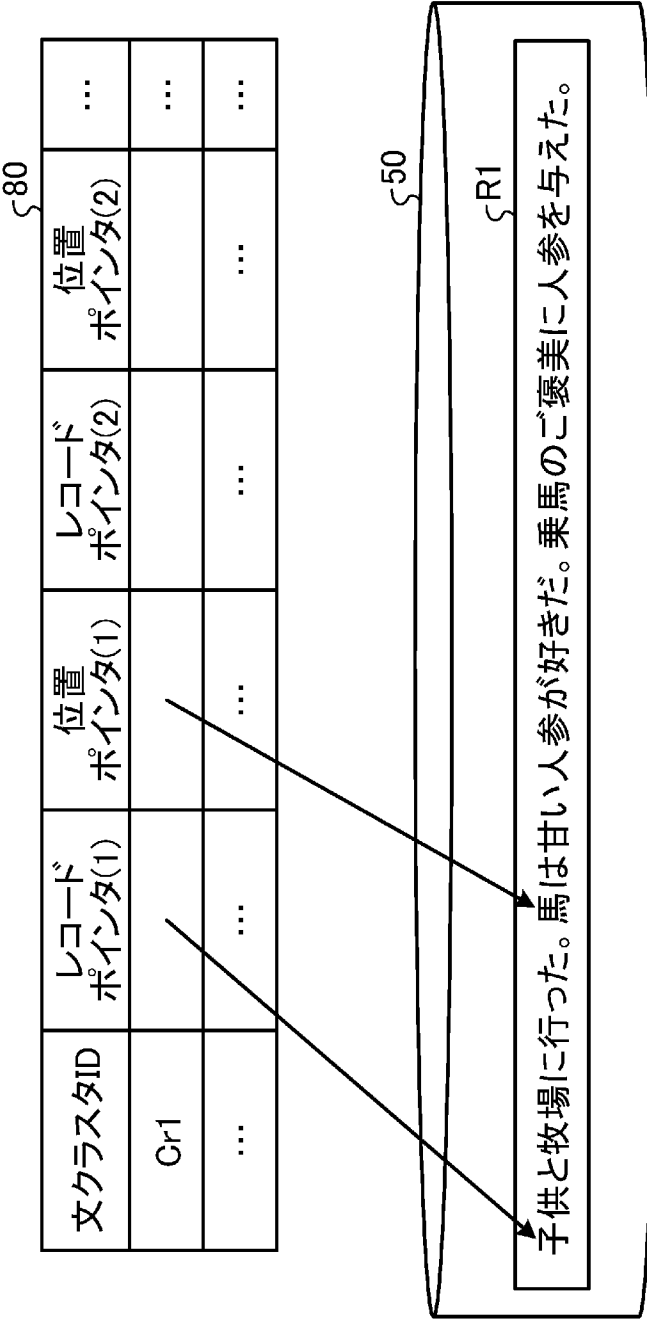
[図4]



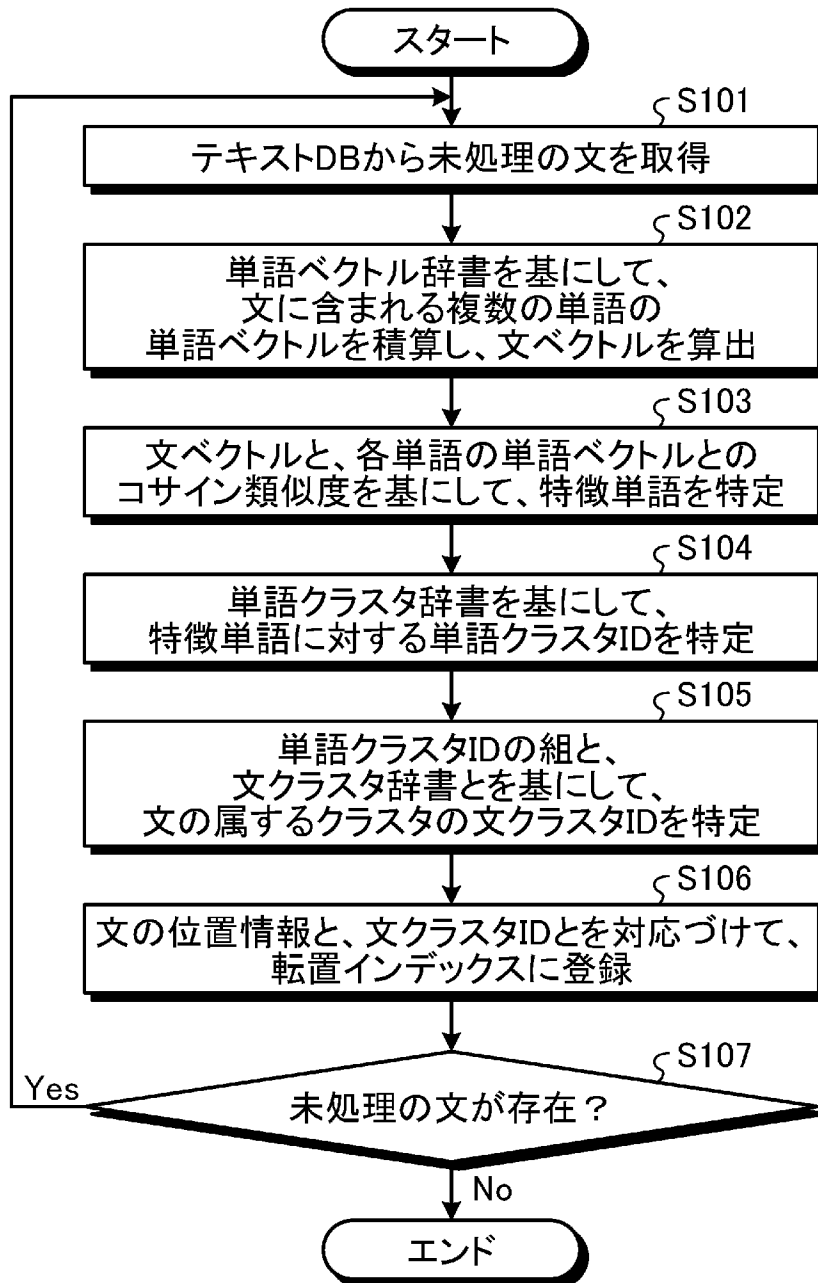
[図5]



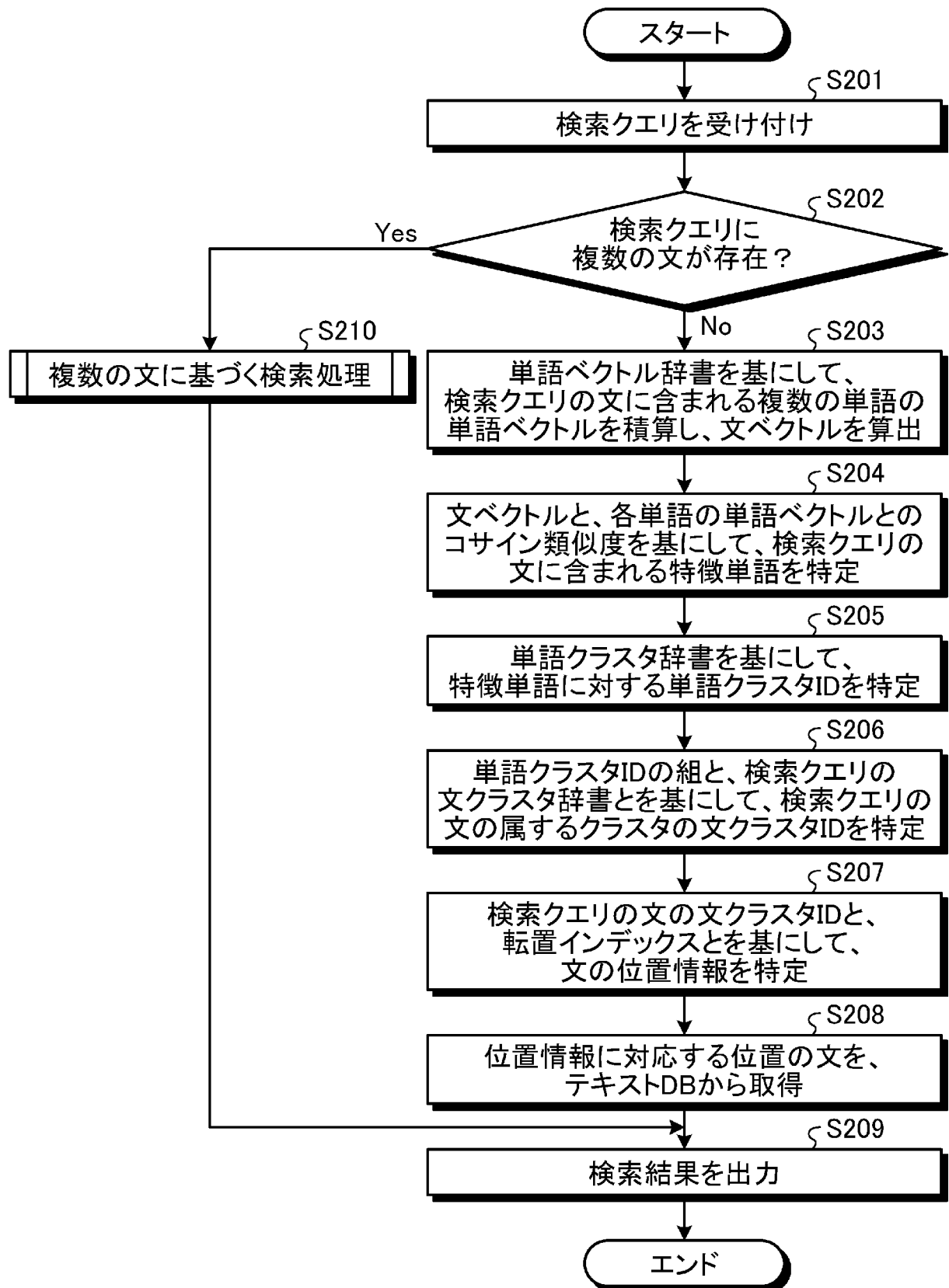
[図7]



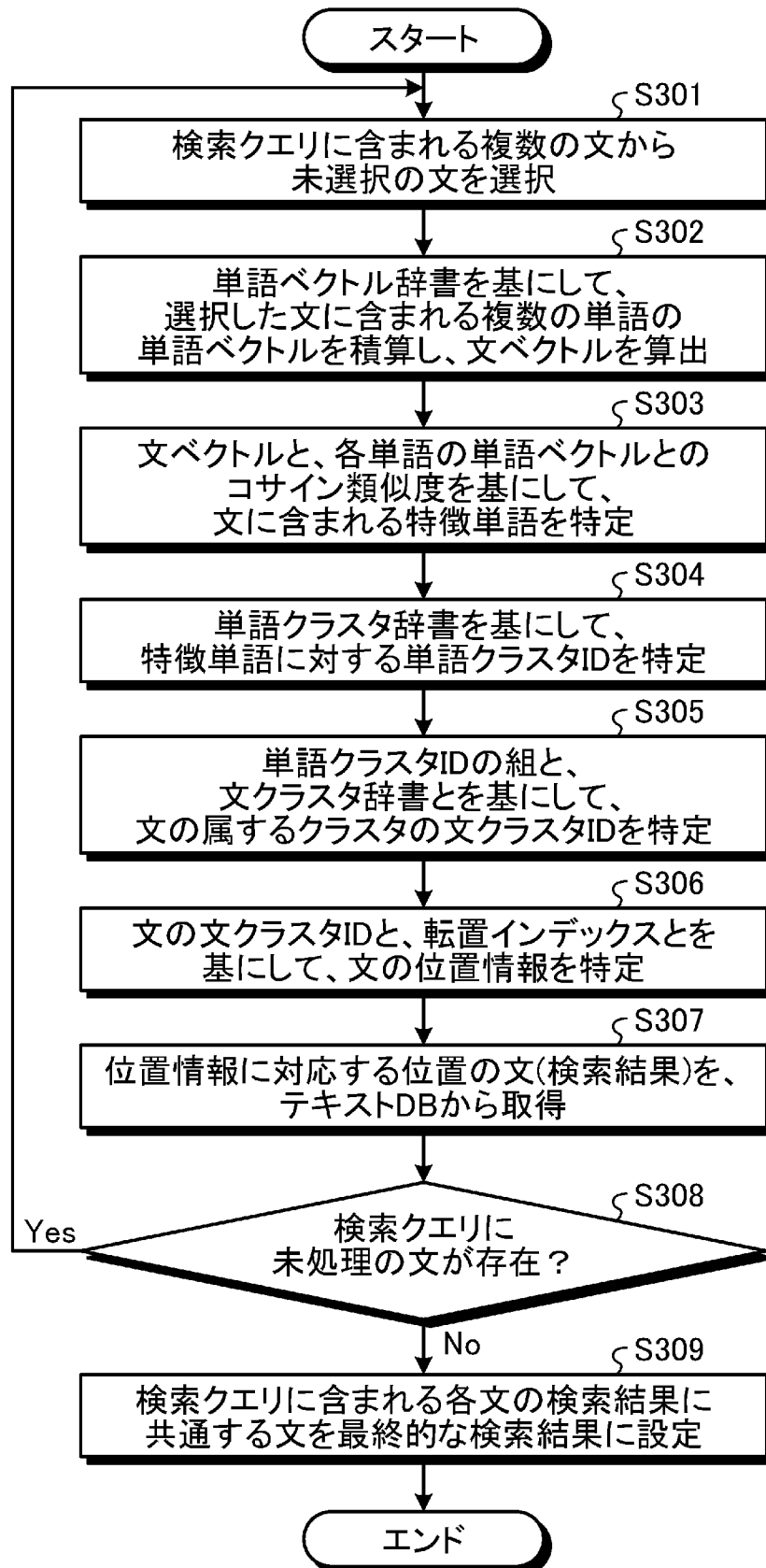
[図8]



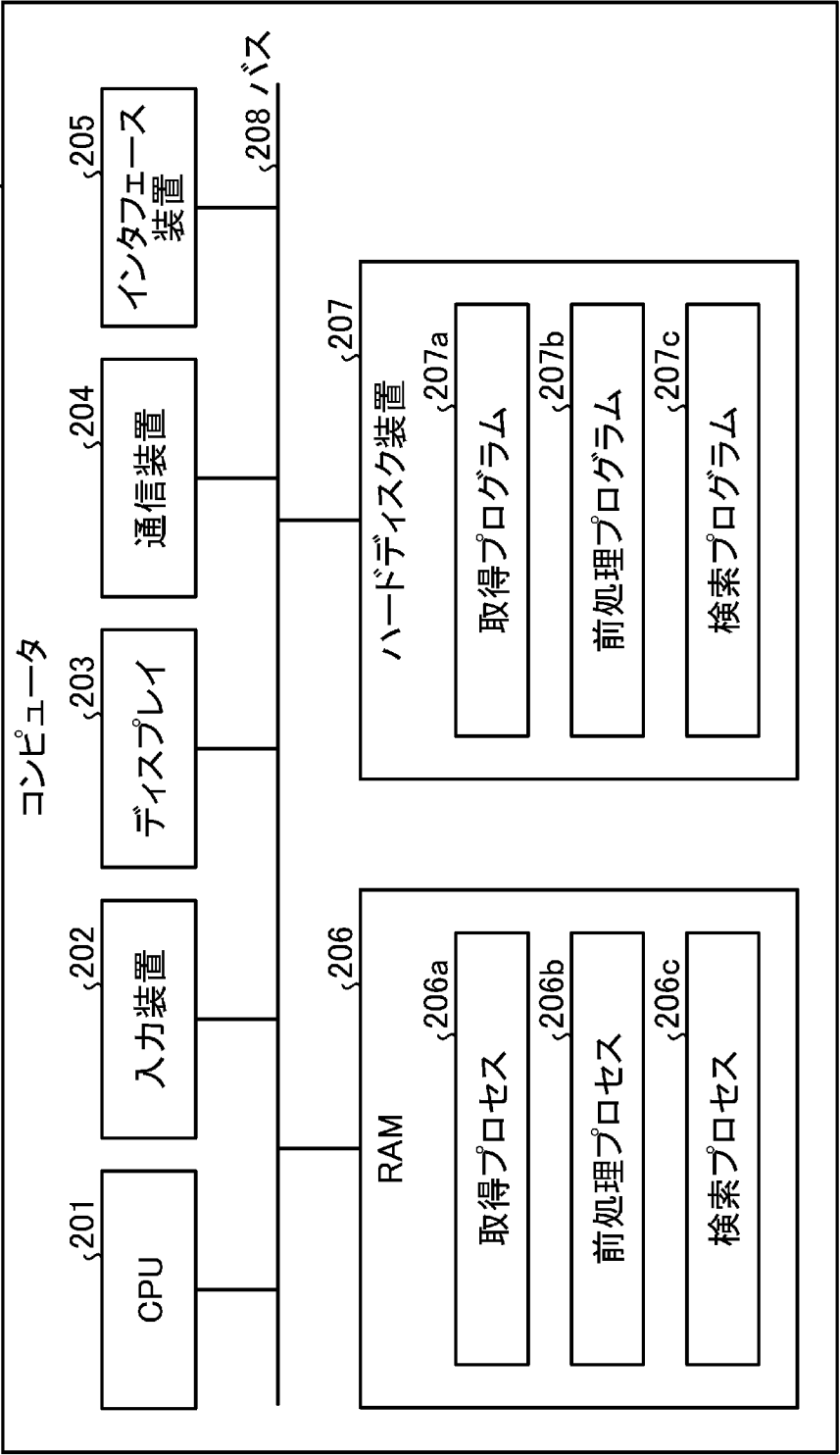
[図9]



[図10]



[図12]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2022/027902

A. CLASSIFICATION OF SUBJECT MATTER**G06F 16/35**(2019.01)i

FI: G06F16/35

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F16/00-16/958

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2022

Registered utility model specifications of Japan 1996-2022

Published registered utility model applications of Japan 1994-2022

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2020-004156 A (FUJITSU LTD) 09 January 2020 (2020-01-09) entire text, all drawings	1-18
A	JP 2019-211884 A (UNIV TOTTORI) 12 December 2019 (2019-12-12) entire text, all drawings	1-18
A	JP 2016-207141 A (YAHOO JAPAN CORP) 08 December 2016 (2016-12-08) entire text, all drawings	1-18



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

07 October 2022

Date of mailing of the international search report

18 October 2022

Name and mailing address of the ISA/JP

Japan Patent Office (ISA/JP)**3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915****Japan**

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2022/027902

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP	2020-004156	A	09 January 2020	US 2020/0004817 A1 entire text, all drawings	
JP	2019-211884	A	12 December 2019	(Family: none)	
JP	2016-207141	A	08 December 2016	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06F 16/35(2019.01)i FI: G06F16/35										
B. 調査を行った分野										
調査を行った最小限資料（国際特許分類（IPC）） G06F16/00-16/958										
最小限資料以外の資料で調査を行った分野に含まれるもの <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922 - 1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971 - 2022年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2022年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2022年</td> </tr> </table>			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2022年	日本国実用新案登録公報	1996 - 2022年	日本国登録実用新案公報	1994 - 2022年
日本国実用新案公報	1922 - 1996年									
日本国公開実用新案公報	1971 - 2022年									
日本国実用新案登録公報	1996 - 2022年									
日本国登録実用新案公報	1994 - 2022年									
国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）										
C. 関連すると認められる文献										
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号								
A	JP 2020-004156 A（富士通株式会社）09.01.2020（2020 - 01 - 09） 全文全文	1-18								
A	JP 2019-211884 A（国立大学法人鳥取大学）12.12.2019（2019 - 12 - 12） 全文全文	1-18								
A	JP 2016-207141 A（ヤフー株式会社）08.12.2016（2016 - 12 - 08） 全文全文	1-18								
☐ C欄の続きにも文献が列挙されている。 ☒ パテントファミリーに関する別紙を参照。										
<table border="0"> <tr> <td> * 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献 </td> <td> “T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献 </td> </tr> </table>			* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献						
* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献									
国際調査を完了した日 07.10.2022	国際調査報告の発送日 18.10.2022									
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 原 秀人 5N 9644 電話番号 03-3581-1101 内線 3586									

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2022/027902

引用文献			公表日	パテントファミリー文献	公表日
JP	2020-004156	A	09.01.2020	US 2020/0004817 A1 全文全図	
JP	2019-211884	A	12.12.2019	(ファミリーなし)	
JP	2016-207141	A	08.12.2016	(ファミリーなし)	