

(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
2020年4月23日(23.04.2020)



(10) 国際公開番号

WO 2020/079751 A1

(51) 国際特許分類:

G06F 17/30 (2006.01)

(21) 国際出願番号: PCT/JP2018/038476

(22) 国際出願日: 2018年10月16日(16.10.2018)

(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: 株式会社島津製作所 (SHIMADZU CORPORATION) [JP/JP]; 〒6048511 京都府京都市中京区西ノ京桑原町1番地 Kyoto (JP).

(72) 発明者: 石川 勇 樹 (ISHIKAWA Yuki); 〒6048511 京都府京都市中京区西ノ京桑原町1番地 株式会社島津製作所内 Kyoto (JP).

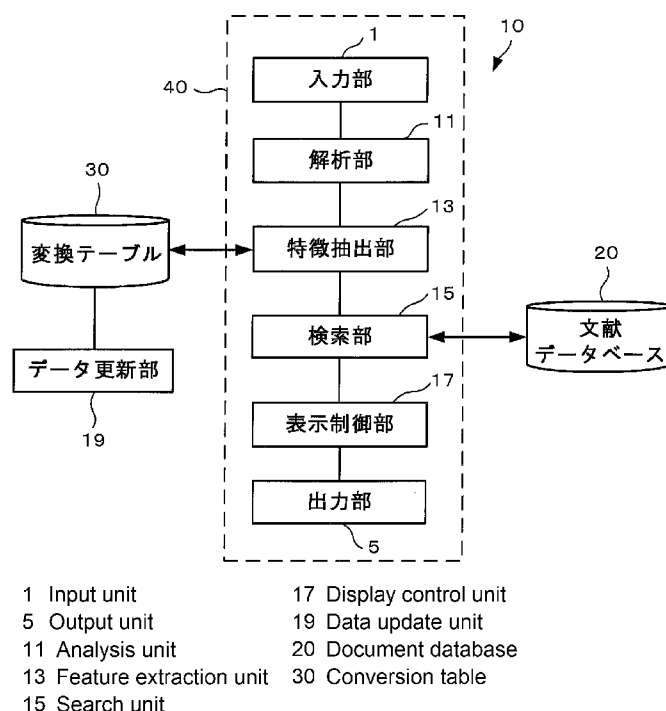
森本 健太郎(MORIMOTO Kentaro); 〒6048511 京都府京都市中京区西ノ京桑原町1番地 株式会社島津製作所内 Kyoto (JP). 大谷 昭成(OTANI Akiyoshi); 〒6048511 京都府京都市中京区西ノ京桑原町1番地 株式会社島津製作所内 Kyoto (JP).

(74) 代理人: 駒井 慎二, 外 (KOMAI Shinji et al.); 〒1010061 東京都千代田区神田三崎町2-6-7 田中衡機ビル3階 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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,

(54) Title: DOCUMENT SEARCH METHOD AND DOCUMENT SEARCH SYSTEM

(54) 発明の名称: 文献検索方法および文献検索システム



(57) Abstract: A document search method and document search system are provided which search documents faster by rendering unnecessary replacement and combination of words in document data as a means of solving the problem of synonyms contained in the document data, and which can provide useful information as search results even when search keywords are inputted that are not present in the search target. To this end, a database 20, by converting words contained in the document data to semantic representation vectors automatically acquired by machine learning, interprets synonyms

HN, HR, HU, ID, IL, IN, IR, IS, 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, 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,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, 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, 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)）

as having a short distance between vectors and renders combination and replacement, and the setting of index words, unnecessary. If a search target contained in a search query is not present in the database 20, a search unit 15 refers to a conversion table 30 which stores similar search targets similar to the search target, and acquires similar search targets similar to said search target. Then, the conversion table 30 is searched with a search query after replacing the search target included in the search query with the acquired similar search target.

(57) 要約：文献データに含まれる類似語を解決するための、文献データ中の単語の置き換えや組み合わせを不要にして、高速な文献検索を行うとともに、検索対象中に存在しない検索キーワードが入力されても、検索結果として有用な情報を提供できる文献検索方法および文献検索システムを提供する。そのため、データベース20は、文献データに含まれる単語を機械学習により自動獲得した意味を表現するベクトルに変換することで、類似語をベクトル間の距離が近いものとして解釈し、組み合わせや置換・索引語の設定を不要とする。検索部15は、検索クエリに含まれる検索対象がデータベース20に存在しない場合、検索対象に類似する類似検索対象が格納された変換テーブル30を参照して、その検索対象と類似する類似検索対象を取得する。そして、検索クエリに含まれる検索対象を、取得した類似検索対象で置換後の検索クエリによって変換テーブル30を検索する。

明 細 書

発明の名称：文献検索方法および文献検索システム

技術分野

[0001] 本発明は、入力された検索クエリに基づいて大量の文献データから検索対象を検索する文献検索方法および文献検索システムに関する。

背景技術

[0002] 近年におけるコンピュータ技術、通信網の発達により、電子化された大量の文書、情報等が蓄積されたデータベースから所望の文書データ、情報等を検索する需要が高まっている。一般的な情報検索システムでは、検索者が検索したい内容を最もよく表している検索語（キーワード）を入力することで、データベースに対する検索処理（索引語とのマッチング処理）が実行され、特定の条件に合致する文書データ等が検索結果として出力される。

[0003] 文献には同意異表示な言葉（例えば、インターフェース、インターフェイス等の翻訳化の相違や送り仮名相違など）や、類似語（例えば、計算機、パソコン、コンピュータなど）が含まれており、従来の検索方法では、これら（以下、類似語と呼ぶ）を同一のものとして処理する工夫が必要であった。例えば、検索対象の類似語をすべての組み合わせで置換したり、複数の索引語を文献に紐付けることで検索漏れを少なくしていた。しかし、組み合わせ数が増えることで検索時間が長くなったり、最適な索引語を設定できず、必ずしも必要な検索ができないという問題があった。同様に、検索クエリの類似語の解決も必要であった。

[0004] 例えば特許文献1は、文献データの登録時に、登録する文献データに固定キーワードを付与して情報検索システムに登録しておき、検索時に、その固定キーワードを指定して短時間で文献データベースを検索する技術を開示している。

[0005] ところで、人工知能分野の自然言語処理では機械学習を応用して、単語そのものを扱うのではなく、単語の意味を表現したベクトルに変換する技術が

用いられている。例えば、大量の文章情報をニューラルネットワークで学習し、ある単語の前後に出現する単語の分布から単語の意味を表すベクトル空間を生成して、それぞれの単語をベクトルに変換する。

[0006] 単語をベクトルに変換することによって、類似語はベクトル間の距離が近いものとして解釈することができる。したがって、検索対象に含まれる単語をベクトルに変換すれば、類似語は組み合わせや置換・索引語の設定が不要となる。機械学習により単語のベクトル空間を獲得する際に、検索対象である文献データを学習データとし、あるいは、その文献データにWeb上で入手できる一般文書などを加えて学習データとすることで単語のベクトル表現は自動で獲得できる。

先行技術文献

特許文献

[0007] 特許文献1：特開2007-200352号公報

発明の概要

発明が解決しようとする課題

[0008] 上述した方法では、検索対象である文献データに含まれるすべての単語のベクトル表現は獲得できるが、検索クエリに含まれる単語は多種多様であるため、検索クエリに含まれる単語があらかじめ生成したベクトル空間に含まれていない場合に対応できない。これは学習データの単語に不足がある場合や新語である場合など、避けられない問題である。ここでは、ベクトルに変換できない単語を未知語と呼ぶ。

[0009] 特許文献1は、種々の検索の語句や表現のバリエーション等を含む類似語を共通登録キーワードとし、文献中に共通登録キーワードを埋め込んだ文献データベースを検索する構成としている。このとき、類似語を共通登録キーワードに変換するため、類似語間の差の情報を消失し、検索結果のリランキングの精度が落ちる点が挙げられる。また、共通登録キーワードの更新時に文献データベース上のすべての文献の更新が必要となるため、計算コストが

高いという問題が挙げられる。

[0010] 本発明は上述した課題、すなわち、検索対象や検索クエリにおける類似語の問題に鑑みてなされたものであり、類似語を含む文献の検索時間を短縮し、検索精度を向上できる文献検索方法および文献検索システムを提供することを目的とする。

[0011] また、本発明は検索クエリに含まれる単語が未知語である場合においても、検索結果として有用な文献情報を提供できる文献検索方法および文献検索システム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

[0012] 上述した課題を解決する一手段として、本発明は、入力された検索クエリに基づいて、文献データベースに格納された検索対象文献を検索する文献検索方法であって、前記文献データベースを機械学習により自動獲得した意味を表現するベクトルに変換する工程と、前記検索クエリに含まれる検索キーワードがベクトルに変換可能か判定する工程と、前記検索クエリに含まれる検索キーワード(単語)がベクトルに変換できないと判定された場合、変換テーブルを用いてベクトル空間に含まれる単語と最も近い単語に変換する工程と、前記変換後の検索クエリをベクトル化する工程と、前記検索クエリによって前記文献データベースを検索する工程と、前記検索で得られた検索結果を表示する工程とを備えることを特徴とする。

[0013] 上記の文献検索方法において、好ましくは、前記文献データベースは、あらかじめ学習によりベクトル化した結果が格納されており、前記変換テーブルには、検索キーワードに類似する類似検索キーワードが格納されている。また、好ましくは、前記類似検索キーワードが複数ある場合、最も類似する類似検索キーワードで前記置換を行う。また、好ましくは、前記検索キーワードに類似する新たな類似検索キーワードを前記変換テーブルに追加して該変換テーブルを更新する工程をさらに備える。また、好ましくは、前記検索クエリに含まれる検索キーワードが前記文献データベースに存在する場合においても前記変換テーブルを参照して該検索キーワードと類似する類似検索

キーワードを取得する工程と、前記検索クエリに含まれる検索キーワードによる検索結果とともに、前記取得した類似検索キーワードによる検索結果を表示する工程とをさらに備える。さらに、好ましくは、前記変換テーブルを表示装置に表示し、指定すべき検索キーワードを選択可能とする工程をさらに備える。また、好ましくは、前記検索キーワードには化合物名、化学式、化学反応式、化学方程式、分析装置名、分析手法のうち少なくとも1つが含まれ、前記文献データベースには前記検索キーワードについての分析レポート、分析関連論文、分析関連特許文献のうち少なくとも1つが含まれる。

発明の効果

[0014] 本発明によれば、文献データに含まれる単語を機械学習により自動獲得した意味を表現するベクトルに変換することで、類似語をベクトル間の距離が近いものとして解釈でき、組み合わせや置換・索引語の設定を不要とすることで検索時間の短縮と、検索精度の向上が実現できる。

[0015] また、本発明によれば、検索クエリに含まれる単語が未知語である場合に、ベクトル空間に含まれる最も近い単語に変換する変換テーブルを用いることで、多種多様な検索クエリに対応し、検索結果として有用な文献情報を提供できる。

図面の簡単な説明

[0016] [図1]本発明の実施形態に係る文献検索システムの構成の一例を示すブロック図である。

[図2]文献検索システムにおける検索対象についての検索処理手順を示すフローチャートである。

[図3]文献検索システムにおいて変換テーブルより取得した複数の検索対象をディスプレイ上に表示する例を示す図である。

[図4]検索範囲を拡張して検索する例を示すフローチャートである。

[図5]文献検索システムをネットワークを介して接続した構成例を示す図である。

発明を実施するための形態

- [0017] 以下、本発明に係る実施形態について添付図面を参照して詳細に説明する。図1は、本発明の実施形態に係る文献検索システムの構成の一例を示すブロック図である。図1の文献検索システム10は、検索者であるユーザ等が入力した検索クエリに基づいて検索対象（事例、文献等）を検索し、その検索結果を出力する検索処理部40を備える。
- [0018] 検索処理部40は、入力部1、解析部11、特徴抽出部13、検索部15、表示制御部17、および出力部5を備える。文献検索システム10はさらに、文献データベース（以下、単にデータベースともいう）20、変換テーブル30、および変換テーブル30に蓄積されたデータを更新するデータ更新部19を備える。
- [0019] 本実施形態に係る文献検索システムにおける検索対象には、例えば、分析レポート、分析関連論文、分析関連特許文献等が含まれる。分析レポートとは、例えば、「試料△△の残留農薬を装置〇〇を使用して分析した。」等の分析事例、分析結果のレポートであり、アプリケーションニュース、アプリケーションノートも分析レポートに含まれる。
- [0020] 文献検索システム10の入力部1には、ユーザ等により、検索対象を文字列、文章等で表した検索クエリ（検索キーワード、検索文字列ともいう。）が入力される。検索クエリは、検索文、検索キーワード等、検索条件を示す文字列等からなり、検索対象が分析レポートであれば、分析関連検索キーワード、分析関連化合物名、分析関連分析対象物名等が含まれる。
- [0021] 入力部1は、例えば、キーボード、マウス等の入力デバイス、ディスプレイ等からなり、ディスプレイ上の入力画面を使用して検索クエリが入力される。ディスプレイは、検索結果等が表示される出力部5でもある。
- [0022] 解析部11は、入力部1より入力された検索クエリに対して、あらかじめ定義された検索辞書をもとに形態素解析を行って、最小単位の単語に分割する。そして、特徴抽出部13は検索クエリを表現するベクトル（単語ベクトル、あるいは特徴ベクトルともいう。）を算出する。例えば、単語ベクトルは、データベース20に蓄積された文献データ、ウェブ（Web）などから

集められる一般データをもとに学習し、獲得しておく。

[0023] 一方、検索クエリの単語から単語ベクトルを算出できない場合には、特徴抽出部13は、変換テーブル30を参照して類似検索対象を抽出し、検索クエリの変換を行ったあと、単語ベクトルに変換する。変換テーブル30は単語ベクトルに含まれる単語と含まれない単語を、別の距離尺度で計算してリスト化したものである。例えば検索クエリとして化合物名を入力する場合、変換テーブルの距離尺度として化合物の類似度や用途、毒性などの物理量が挙げられる。

[0024] 検索部15は、特徴抽出部13より取得した検索キーワードの単語ベクトルをもとに、データベース20に蓄積された文献データを検索し、それを表示制御部17に出力する。

[0025] 出力部5は、表示制御部17における表示内容、表示形式等の制御結果に応じた情報を表示する。これによりユーザ等は、所望の検索対象あるいはそれと類似する検索対象を取得できる。

[0026] なお、データ更新部19は、変換テーブル30にない新語、その類義語、同義語等を新たに変換テーブル30に登録するという継続的なメンテナンスを行うことで、最新の検索対象が蓄積された状態を維持する。

[0027] 文献検索システム10において、データベース20とは別に変換テーブル30を備えるのは、データベース20は機械学習による単語ベクトル獲得の再学習が必要となるのに対して、変換テーブル30は蓄積されたデータの更新が容易である点、データの蓄積構造が簡単である点等の理由による。

[0028] 次に、本実施形態の文献検索システムにおける検索処理手順について説明する。図2は、本実施形態の文献検索システムにおける検索対象についての検索処理手順を示すフローチャートである。図2のステップS11で、ユーザ等により検索クエリ（検索キーワード）が入力される。次のステップS13で、形態素解析により検索クエリを最小単位の形態素に分割し、単語ベクトルを算出する。

[0029] ステップS15では、文献データ（例えば、分析レポート、分析関連論文

、分析関連特許文献等）をベクトル化したデータが蓄積されたデータベース 20 を参照する検索処理を行う。この検索処理では、検索クエリ（検索キーワード）に関連のある検索対象を検索し、それを検索結果とする。

[0030] ステップ S 17 では、ステップ S 15 での検索結果よりデータベース 20 中に検索対象が存在するか否か、すなわち、検索対象が「未知語」からなる検索対象か否かを判定する。例えば、検索クエリに検索対象として化合物等の名称が含まれているが、その化合物名がデータベース 20 中に存在しなければ、その化合物名は「未知語」と認識される。

[0031] データベース 20 中に検索対象が存在すれば、ステップ S 27 において、検索結果を表示する。一方、データベース 20 に含まれない検索クエリが入力され、検索対象が未知語であってデータベース 20 中にその検索対象が存在しない場合には、ステップ S 19 において変換テーブル 30 を参照する。

[0032] 変換テーブル 30 には、上述したように検索対象毎に類似する検索対象との関連を計算し、それらがリスト化されている。例えば、変換テーブル 30 が化合物変換テーブルの場合には、既知の化合物ごとに、その化合物と類似する化合物との関連を計算した結果が蓄積されている。

[0033] ステップ S 21 において検索クエリに類似する検索対象を取得する。より具体的には、検索クエリが化合物であり、その化合物としてフタル酸エステルが入力された場合、変換テーブル 30 中においてリスト化された、フタル酸エステルに連なるフタル酸ジプロピル、フタル酸ジネブチル、フタル酸ジノニル、フタル酸ベンチルイソベンチル、フタル酸ジメチル等の類似化合物を検索クエリとする。

[0034] よって、ステップ S 21 における処理では、入力された検索クエリを、類似した検索クエリに置換して変換テーブル 30 を検索していることになる。検索クエリの置換は、取得された類似化合物が複数ある場合、例えば、その中から最も類似する化合物によって行う。

[0035] ステップ S 23 では、ステップ S 21 で取得された複数の類似化合物の中から、例えば、検索クエリの特徴等を考慮した最適な類似化合物を選択する

。

[0036] なお、変換テーブル30中の検索対象には、化合物のみならず、化学式、分析装置名、化学的な方程式、分析手法等を含めてもよい。また、検索対象に類似する検索対象を取得する際の類似するか否かの判定の精度を上げるために、例えば、構造式や極性等の追加入力をする構成としてもよい。

[0037] ステップS25において、上記のステップS23で選択された類似化合物を変換テーブル30より抽出する等の検索処理を行い、続くステップS27において、検索結果を表示する。

[0038] 図3は、本実施形態の文献検索システムにおいて、変換テーブル30より取得した複数の検索対象をディスプレイ上に表示する例である。ここでは、検索対象物が化合物Aであって、変換テーブル30より複数の類似化合物が取得された場合、ディスプレイ41上にそれら複数の類似化合物A1, A2, A3を表示し、ユーザが、それら類似化合物に対応して設けたボックス43aにチェックを入れることで、検索対象（類似化合物）を選択できる構成となっている。

[0039] 検索対象物である他の化合物B, Cについても、化合物Aと同様、それぞれの類似化合物B1～B4, C1が表示され、ボックス43b, 43cで選択できる構成となっている。

[0040] 本実施形態の文献検索システムにおける検索対象についての検索処理は、図2に示す例に限定されず、以下のように検索範囲の拡張を伴う検索を行ってもよい。

[0041] 図4は、検索範囲を拡張して検索する例を示すフローチャートである。なお、図4において、図2に示すステップと同一のステップには同一の符号を付し、それらの説明を省略する。

[0042] 図4に示す検索処理では、ステップS15での検索結果より、ステップS17においてデータベース20中に検索対象が存在しないと判断された場合、ステップS19, S21, S23の処理を実行するが、ここでは、データベース20中に検索対象が存在する場合においても、ステップS19, S2

1, S 2 3と同様の処理を行う。

[0043] すなわち、ステップS 1 7での判断結果がN Oで、データベース2 0中に検索対象が存在しても、ステップS 3 1において、検索対象毎に類似する検索対象との関連を計算してリスト化された変換テーブル3 0を参照する。

[0044] ステップS 3 3において、検索対象に類似する検索対象を取得する。その際、入力された検索クエリを、類似した検索クエリに置換して変換テーブル3 0を検索する。検索クエリの置換は、取得された複数の類似化合物の中から最も類似する化合物によって行う。

[0045] ステップS 3 5では、検索された検索対象に類似する検索対象の中から最適な類似検索対象を選択する。その結果、ユーザが入力した検索対象の検索結果に追加して、それと類似する検索対象についての検索結果を取得して、それらを表示するという、検索範囲を拡張した検索の結果を得ることができる。

[0046] 図5は、本実施形態の文献検索システムをネットワーク（公衆通信網）を介して接続した構成例を示している。図5に示すように本実施形態に係る文献検索システム1 0は、例えばインターネット環境において、ユーザ等からの要求に応じて事例の検索を可能にし、検索結果をユーザ等に提供する。

[0047] そのため、インターネット等の情報通信ネットワーク7 0を介して、文献検索システム1 0と、複数のユーザ端末6 0 a…6 0 nとが通信可能に接続されている。文献検索システム1 0内の通信部6 1は、ネットワーク7 0とのインターフェイスである。制御部6 5は、検索処理部4 0を含む文献検索システム1 0の全体の制御を司る、例えばマイクロプロセッサ等で構成される。

[0048] メモリ6 7には、上述した検索処理部4 0における検索対象の検索処理プログラム等が格納されている。制御部6 5は、これらのプログラムを読み出して、図2および図4に示す所定の処理等を実行する。なお、文献検索システム1 0全体を、ネットワーク7 0に接続されたサーバ装置と位置づけることもできる。

- [0049] 以上説明したように本実施形態によれば、文献データを格納したデータベース、検索対象と類似する類似検索対象を格納した変換テーブルを用意し、検索クエリとして未知語が入力された場合でも、変換テーブルを参照して類似検索対象を取得し、それを検索結果として表示することで、ユーザに有用な情報を確実に提示できる。
- [0050] よって、学習データが少量の場合に未知語が含まれることを許容し、未知語が入力された場合であっても、検索対象（化合物）の既知の関係性を利用した類似検索対象（類似化合物）の選択を行い、その類似検索対象（類似化合物）による検索を行うことで検索精度を向上できる。
- [0051] また、データベースとは別に変換テーブルを備えることで、データの蓄積構造が簡単な変換テーブルに蓄積されたデータの更新が容易となる。これにより、データベースに類似語情報を持たせた場合のように、計算コストが大きいデータベースの更新作業を回避できる。
- [0052] さらに、新語、その類義語、同義語等を新たに継続的に変換テーブルに登録することで、変換テーブルを最新の検索対象が蓄積された状態に維持できる。
- [0053] また、ユーザが入力した検索クエリ（化合物）が未知語でない場合であっても、変換テーブルによる類似化合物の検索を行い、ユーザが入力した検索クエリ（化合物）の検索結果に追加して、類似検索対象（類似化合物）の検索結果を提示できる。

符号の説明

- [0054] 1 入力部
5 出力部
10 文献検索システム
11 解析部
13 特徴抽出部
15 検索部
17 表示制御部

- 19 データ更新部
- 20 データベース
- 30 変換テーブル
- 40 検索処理部
- 41 ディスプレイ
- 60a...60n ユーザ端末
- 61 通信部
- 65 制御部
- 67 メモリ
- 70 情報通信ネットワーク

請求の範囲

- [請求項1] 入力された検索クエリに基づいて、データベースに格納された文献データを検索する文献検索方法であって、
- 前記検索クエリに含まれる検索対象が前記データベースに存在するか否かを判定する工程と、
- 前記データベースに前記検索対象が存在しないと判定された場合、変換テーブルを参照して該検索対象と類似する類似検索対象を取得する工程と、
- 前記検索クエリに含まれる検索対象を前記取得した類似検索対象で置換する工程と、
- 前記置換後の検索クエリによって前記データベースを検索する工程と、
- 前記検索で得られた検索結果を表示する工程と、
- を備えることを特徴とする文献検索方法。
- [請求項2] 前記データベースには、機械学習によって単語の意味を表現するベクトルが獲得されて、文献データをベクトル空間に写像したデータが格納されており、前記変換テーブルには、既知の情報に基づく検索対象に類似する類似検索対象が格納されていることを特徴とする請求項1に記載の文献検索方法。
- [請求項3] 前記類似検索対象が複数ある場合、最も類似する類似検索対象で前記置換を行うことを特徴とする請求項1に記載の文献検索方法。
- [請求項4] 前記検索対象に類似する新たな類似検索対象を前記変換テーブルに追加して該変換テーブルを更新する工程をさらに備えることを特徴とする請求項1に記載の文献検索方法。
- [請求項5] 前記検索クエリに含まれる検索対象が前記データベースに存在する場合においても前記変換テーブルを参照して該検索対象と類似する類似検索対象を取得する工程と、
- 前記検索クエリに含まれる検索対象による検索結果とともに、前記

取得した類似検索対象による検索結果を表示する工程と、
をさらに備えることを特徴とする請求項 1 に記載の文献検索方法。

[請求項6] 前記検索クエリには化合物名、化学式、化学反応式、化学方程式、分析装置名、分析手法のうち少なくとも 1 つが含まれ、前記文献データには前記検索対象についての分析レポート、分析関連論文、分析関連特許文献のうち少なくとも 1 つが含まれることを特徴とする請求項 1 ～ 5 のいずれか 1 項に記載の文献検索方法。

[請求項7] 入力された検索クエリに基づいて、データベースに格納された文献データを検索する文献検索システムであって、

前記検索クエリに含まれる検索対象が前記データベースに存在するか否かを判定する手段と、

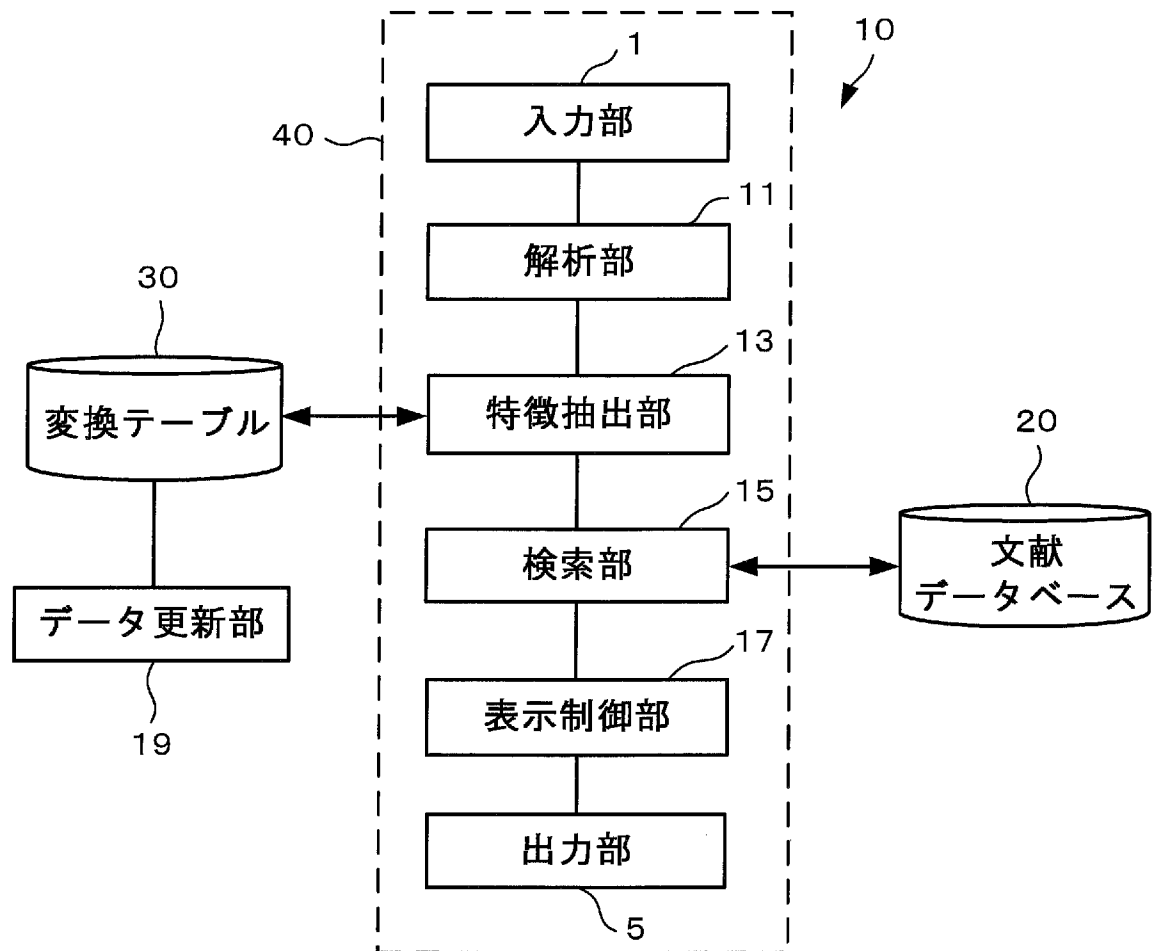
前記データベースに前記検索対象が存在しないと判定された場合、変換テーブルを参照して該検索対象と類似する類似検索対象を取得する手段と、

前記検索クエリに含まれる検索対象を前記取得した類似検索対象で置換する手段と、

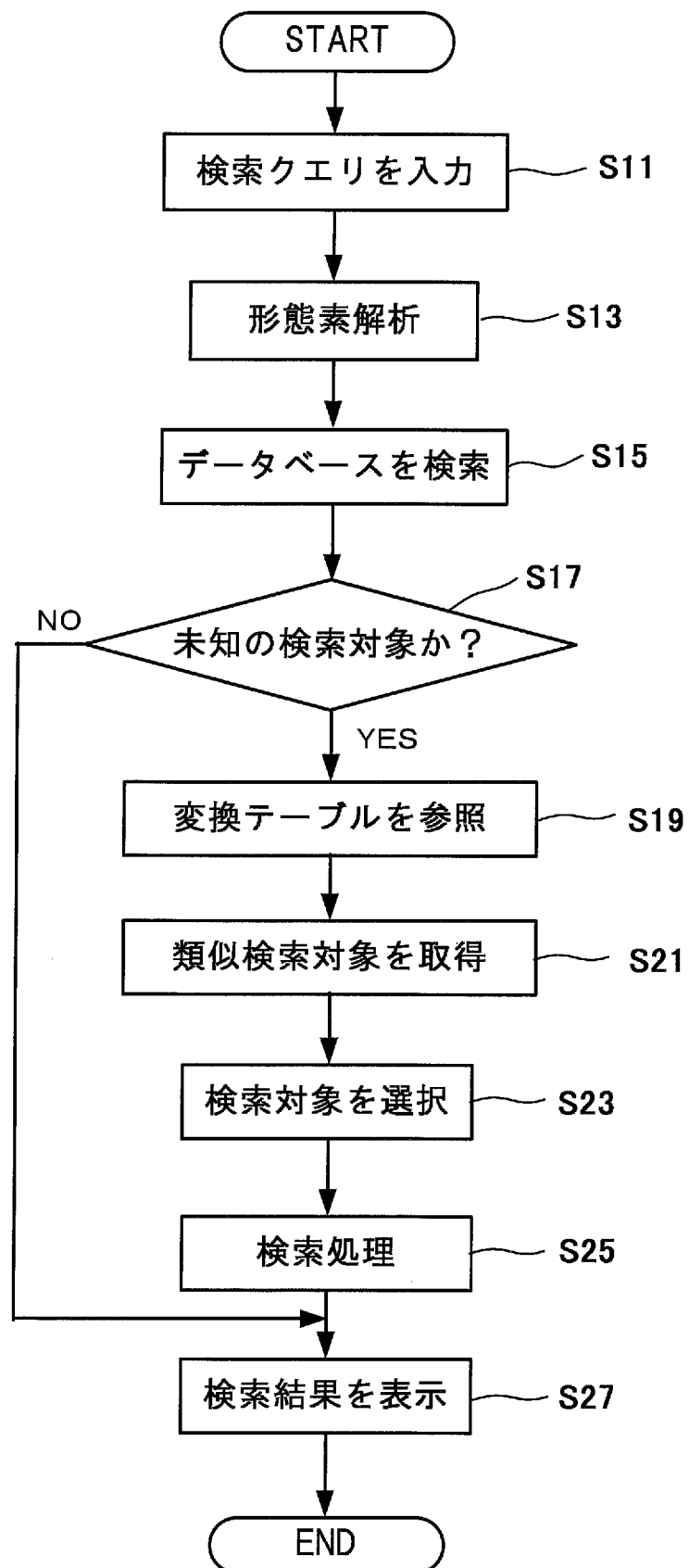
前記置換後の検索クエリによって前記データベースを検索する手段と、

前記検索で得られた検索結果を表示する手段と、
を備えることを特徴とする文献検索システム。

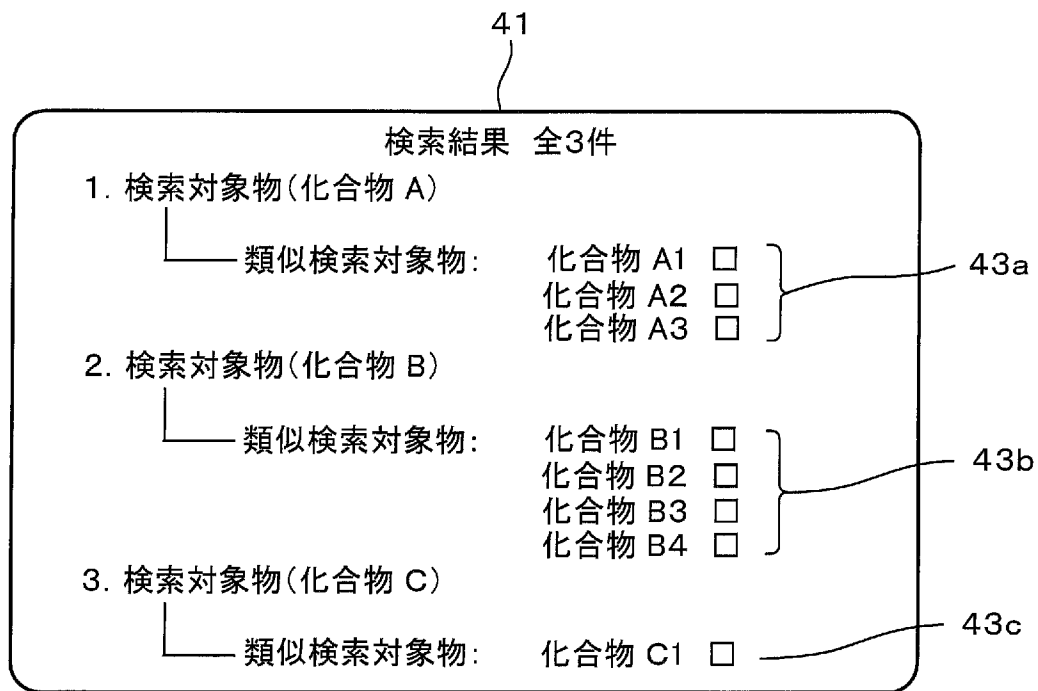
[図1]



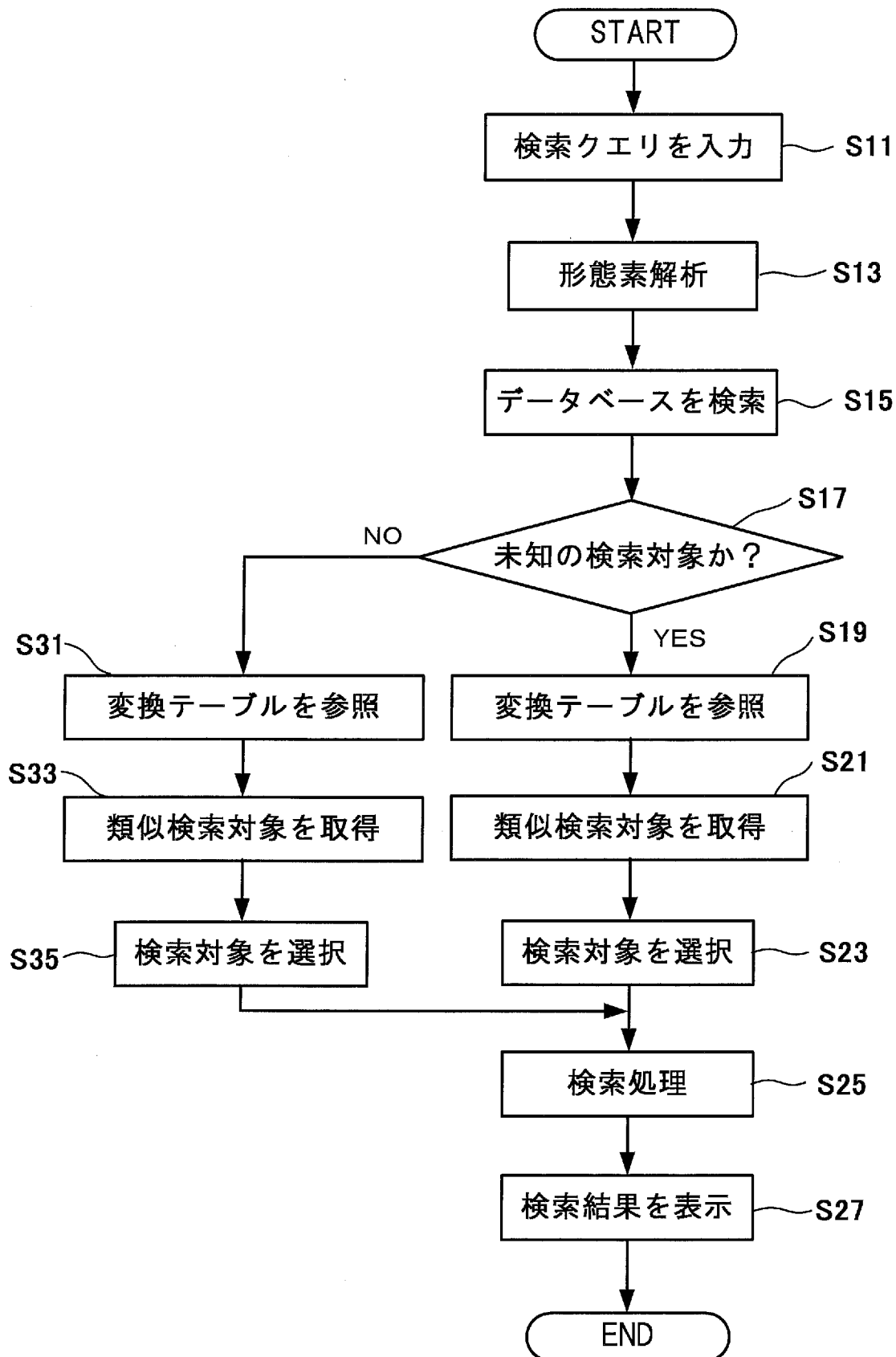
[図2]



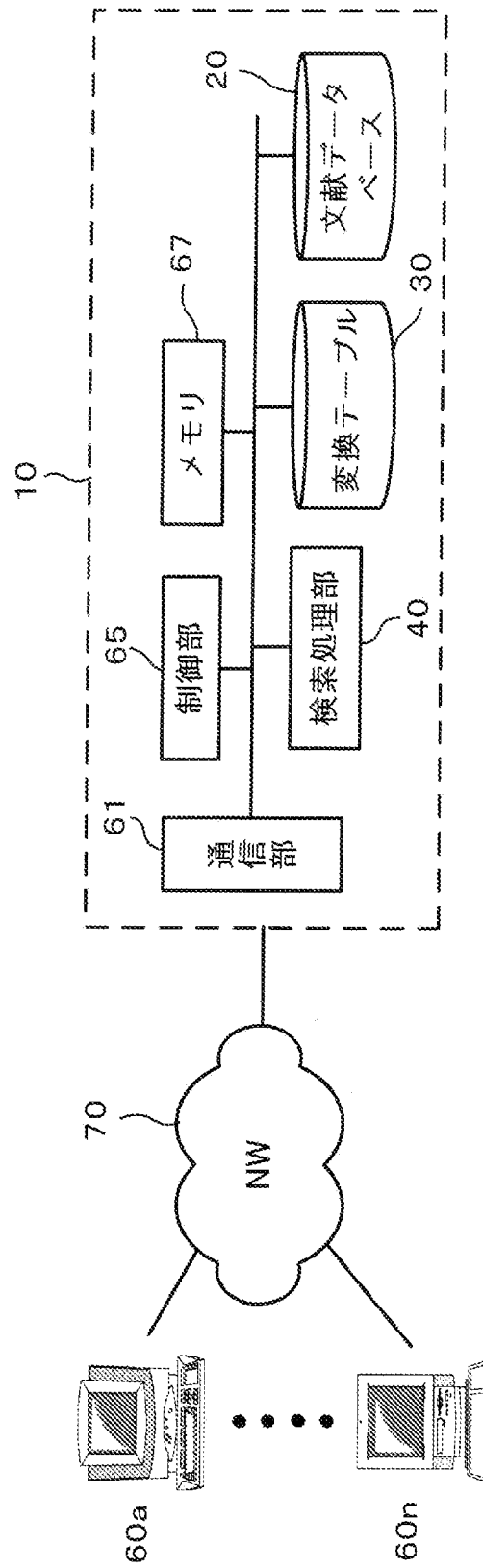
[図3]



[図4]



[図5]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2018/038476

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G06F17/30 (2006.01) i

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)

Int.Cl. G06F17/30

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 1971-2018

Registered utility model specifications of Japan 1996-2018

Published registered utility model applications of Japan 1994-2018

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.
X	JP 2014-153744 A (JAPAN BROADCASTING CORPORATION)	1-4, 7
Y	25 August 2014, paragraphs [0010], [0019]-[0026], [0044]-[0054], [0074]-[0083], fig. 5, 6 (Family: none)	5-6
Y	JP 2002-175329 A (CANON INC.) 21 June 2002, paragraphs [0091]-[0097], [0099], [0100] (Family: none)	5-6



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

05.11.2018

Date of mailing of the international search report

13.11.2018

Name and mailing address of the ISA/

Japan Patent Office

3-4-3, Kasumigaseki, Chiyoda-ku,

Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2018/038476

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2006-163953 A (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 22 June 2006, entire text, all drawings (Family: none)	1-7
A	"Word2Vec how to treat words that aren't in the vocabulary", Data Science Stack Exchange [online], 28 September 2016, [retrieved on 05 November 2018], Internet <URL:https://datascience.stackexchange.com/questions/13722/word2vec-how-to-treat-words-that-arent-in-the-vocabulary>	1-7

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. G06F17/30(2006.01)i

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))

Int.Cl. G06F17/30

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

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2018年
日本国実用新案登録公報	1996-2018年
日本国登録実用新案公報	1994-2018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
X Y	JP 2014-153744 A (日本放送協会) 2014.08.25, [0010], [0019]-[0026], [0044]-[0054], [0074]-[0083], [図5]-[図6] (ファミリーなし)	1-4, 7 5-6
Y	JP 2002-175329 A (キヤノン株式会社) 2002.06.21, [0091]-[0097], [0099]-[0100] (ファミリーなし)	5-6

☒ 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。

* 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矛盾するものではなく、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

05.11.2018

国際調査報告の発送日

13.11.2018

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)

郵便番号 100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

後藤 昂彦

5M

6291

電話番号 03-3581-1101 内線 3599

C (続き) . 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 2006-163953 A (日本電信電話株式会社) 2006.06.22, 全文, 全図 (ファミリーなし)	1-7
A	"Word2Vec how to treat words that aren't in the vocabulary", Data Science Stack Exchange [online], 2016.09.28, [検索日: 2018.11.05], インターネット<URL: https://datascience.stackexchange.com/questions/13722/word2v ec-how-to-treat-words-that-arent-in-the-vocabulary >	1-7