

(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
2014 年 6 月 26 日(26.06.2014)



(10) 国際公開番号

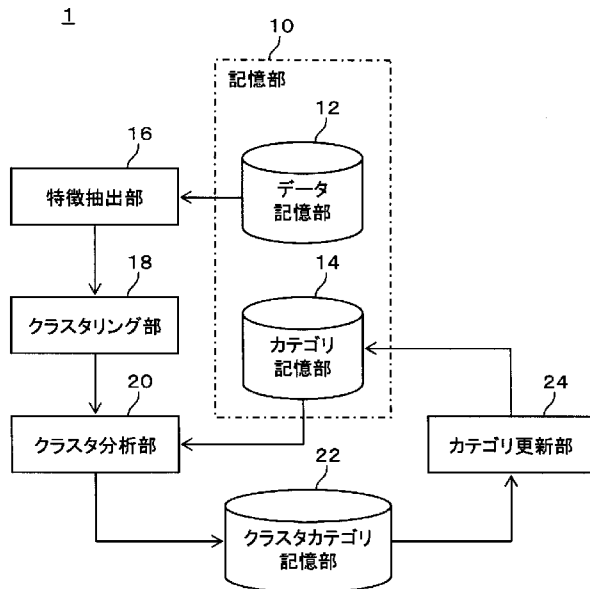
WO 2014/097670 A1

- (51) 国際特許分類:
G06F 17/30 (2006.01)
- (21) 国際出願番号: PCT/JP2013/068852
- (22) 国際出願日: 2013 年 7 月 10 日(10.07.2013)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2012-279624 2012 年 12 月 21 日(21.12.2012) JP
- (71) 出願人: 富士ゼロックス株式会社(FUJI XEROX CO., LTD.) [JP/JP]; 〒1070052 東京都港区赤坂九丁目 7 番 3 号 Tokyo (JP).
- (72) 発明者: 服部 圭悟(HATTORI KEIGO); 〒2200012 神奈川県横浜市西区みなとみらい六丁目 1 番 富士ゼロックス株式会社内 Kanagawa (JP). 増市 博(MASUICHI HIROSHI); 〒2200012 神奈川県横浜市西区みなとみらい六丁目 1 番 富士ゼロックス株式会社内 Kanagawa (JP).
- (74) 代理人: 深澤 潔(FUKASAWA Kiyoshi); 〒1830033 東京都府中市分梅町 2 丁目 3 7 番地 4 - 2 0 2 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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, 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, 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,

[続葉有]

(54) Title: DOCUMENT CLASSIFICATION DEVICE AND PROGRAM

(54) 発明の名称: 文書分類装置及びプログラム



- 10 Storage unit
12 Data storage unit
14 Category storage unit
16 Feature extraction unit
18 Clustering unit
20 Cluster analysis unit
22 Cluster category storage unit
24 Category update unit

(57) Abstract: In order for document data to be assigned with a category expressing the ambiguity in definitions of categories which are pre-assigned to the document data when said ambiguity is present, a feature extraction unit (16) extracts feature information from document data which have been classified into specific categories in advance. A clustering unit (18) sorts the document data into clusters on the basis of the feature information. A category update unit (24) also assigns, to document data which is classified into a given cluster, categories of other document data which is classified into the given cluster as categories of the document data.

(57) 要約: 予め文書データに付与されたカテゴリの定義に曖昧性がある場合であっても、その曖昧性を表現したカテゴリを文書データに付与するために、特徴抽出部 16 は、予め特定のカテゴリに分類された文書データから特徴情報を抽出する。クラスタリング部 18 は、特徴情報に基づいて文書データをクラスタ分けする。カテゴリ更新部 24 は、同一クラスタに分類された文書データに、同一クラスタに分類された他の文書データのカテゴリも自己のカテゴリとして付与する。

WO 2014/097670 A1



NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI 添付公開書類:

(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, — 国際調査報告 (条約第 21 条(3))
MR, NE, SN, TD, TG).

明 細 書

発明の名称：文書分類装置及びプログラム

技術分野

[0001] 本発明は、文書データを分類する文書分類装置及びプログラムに関する。

背景技術

[0002] 文書データを特定のカテゴリに分類する技術が知られている。

[0003] 例えば、特許文献1には、予め特定のカテゴリに分類された学習用データがある場合、カテゴリの特徴と未分類のデータとを比較することで、未分類のデータのカテゴリを決定する方法が開示されている。

[0004] また、特許文献2には、各文書を文書の特徴量に基づいてクラスタ分けし、各クラスタの特徴語を当該クラスタに属する文書のカテゴリとして用いる方法が開示されている。

先行技術文献

特許文献

[0005] 特許文献1：特開平11-167581号公報

特許文献2：特開2005-182611号公報

発明の概要

発明が解決しようとする課題

[0006] ところで、カテゴリ間の境界が曖昧で、文書データを特定のカテゴリに明確に分類できない場合がある。例えば、文書データのカテゴリが「概要」であるとともに「経歴」でもある場合や、「概要」であるとともに「構成」でもある場合がある。また、人が文書データを分類する場合、その人によって分類の仕方が異なるため、文書データに付与されるカテゴリが人によって異なる場合がある。以上のように、カテゴリの独立性の保障が困難であり、また、人に依存してカテゴリ分類にゆらぎが生じるため、予め特定のカテゴリに分類された学習用データにおいては、学習用データとカテゴリとの組み合わせが正確でない場合がある。カテゴリ分類のゆらぎが生じないようにカテ

ゴリを細かく定義することも考えられるが、カテゴリを細かく定義するほどカテゴリの設定コストが増大し、カテゴリを付与する人に依存してカテゴリ分類にゆらぎが生じるおそれがある。また、自動でカテゴリを決定する場合、例えば特定のデータ集合に「概要」という用語が出現しない場合、そのデータ集合が「概要」について表している場合でも、「概要」というカテゴリに分類することはできない。また、自動でカテゴリを決定する場合、分類されたカテゴリの意味を人が解釈できない場合がある。

[0007] 本発明の目的は、予め文書データに付与されたカテゴリの定義に曖昧性がある場合であっても、その曖昧性を表現したカテゴリを文書データに付与することが可能な文書分類装置及びプログラムを提供することである。

課題を解決するための手段

[0008] 請求項 1 に記載の発明は、予め特定のカテゴリに分類された複数の文書データのそれぞれから特徴情報を抽出する特徴抽出手段と、前記特徴情報の出現頻度が類似する文書データを同一クラスタに分類するクラスタリング手段と、同一クラスタに分類された文書データに、前記同一クラスタに分類された他の文書データのカテゴリも自己のカテゴリとして付与するカテゴリ更新手段と、を有することを特徴とする文書分類装置である。

[0009] 請求項 2 に記載の発明は、請求項 1 に記載の文書分類装置であって、前記特徴抽出手段は、前記特定のカテゴリに関連する情報を特徴情報として前記複数の文書データのそれぞれから抽出する、ことを特徴とする。

[0010] 請求項 3 に記載の発明は、請求項 1 又は請求項 2 に記載の文書分類装置であって、前記カテゴリ更新手段は、前記同一クラスタにおける各文書データのカテゴリの出現割合を、前記同一クラスタに分類された文書データに付与する、ことを特徴とする。

[0011] 請求項 4 に記載の発明は、請求項 1 から請求項 3 のいずれか一項に記載の文書分類装置であって、検索クエリを受け、前記検索クエリにカテゴリに関する情報が含まれている場合、文書データに予め付与されたカテゴリ及び前記カテゴリ更新手段によって付与されたカテゴリも検索対象とし、前記検索

クエリを用いて検索し、検索結果を出力する検索手段を更に有する、ことを特徴とする。

[0012] 請求項 5 に記載の発明は、請求項 4 に記載の文書分類装置であって、前記検索手段は、前記同一クラスタにおける各文書データのカテゴリの出現割合に応じた検索結果を出力する、ことを特徴とする。

[0013] 請求項 6 に記載の発明は、請求項 4 又は請求項 5 に記載の文書分類装置であって、前記検索クエリによって検索された文書データのカテゴリと前記検索クエリに含まれるカテゴリに関する情報とが一致しない場合、前記検索された文書データのカテゴリを、前記検索クエリに含まれるカテゴリに変更する更新手段を更に有する、ことを特徴とする。

[0014] 請求項 7 に記載の発明は、請求項 1 から請求項 6 のいずれか一項に記載の文書分類装置であって、前記カテゴリ更新手段によってカテゴリが付与された文書データに基づいて作成された識別器を用いて未分類の文書データにカテゴリを付与するカテゴリ分類手段を更に有する、ことを特徴とする。

[0015] 請求項 8 に記載の発明は、請求項 7 に記載の文書分類装置であって、前記カテゴリ分類手段は、出現割合が最大のカテゴリを前記未分類の文書データに付与する、ことを特徴とする。

[0016] 請求項 9 に記載の発明は、請求項 7 又は請求項 8 に記載の文書分類装置であって、前記カテゴリ分類手段によって処理された前記未分類の文書データを、前記特徴抽出手段、前記クラスタリング手段及び前記カテゴリ更新手段の処理対象の文書データに追加する手段を更に有する、ことを特徴とする。

[0017] 請求項 10 に記載の発明は、請求項 1 から請求項 9 のいずれか一項に記載の文書分類装置であって、前記特徴抽出手段は、文書データにおける特徴情報の出現頻度に応じて、前記クラスタリング手段によるクラスタ分けに用いられる特徴情報を選択する、ことを特徴とする。

[0018] 請求項 11 に記載の発明は、請求項 1 から請求項 10 のいずれか一項に記載の文書分類装置であって、前記カテゴリ更新手段は、前記同一カテゴリにおける各文書データのカテゴリの出現割合に応じてカテゴリを選択して文書

データに付与する、ことを特徴とする。

- [0019] 請求項 1 2 に記載の発明は、コンピュータに、予め特定のカテゴリに分類された複数の文書データのそれぞれから特徴情報を抽出するステップと、前記特徴情報の出現頻度が類似する文書データを同一クラスタに分類するステップと、同一クラスタに分類された文書データに、前記同一クラスタに分類された他の文書データのカテゴリも自己のカテゴリとして付与するステップと、を実行させることを特徴とするプログラムである。

発明の効果

- [0020] 請求項 1, 1 2 に記載の発明によると、カテゴリの定義に曖昧性がある場合であっても、その曖昧性を表現したカテゴリを文書データに付与することが可能となる。
- [0021] 請求項 2 に記載の発明によると、本発明の構成を備えていない場合と比べて、文書分類の精度が向上する。
- [0022] 請求項 3 に記載の発明によると、カテゴリの特徴の割合を文書データに付与することが可能となる。
- [0023] 請求項 4 に記載の発明によると、カテゴリを考慮した検索結果が得られる。
- [0024] 請求項 5 に記載の発明によると、カテゴリの割合を考慮した検索結果が得られる。
- [0025] 請求項 6 に記載の発明によると、本発明の構成を備えていない場合と比べて、文書データのカテゴリをより正確に修正することが可能となる。
- [0026] 請求項 7 に記載の発明によると、カテゴリの曖昧性を表現したカテゴリを未分類の文書データに付与することが可能となる。
- [0027] 請求項 8 に記載の発明によると、本発明の構成を備えていない場合と比べて、より正確なカテゴリを未分類の文書データに付与することが可能となる。
- [0028] 請求項 9 に記載の発明によると、未分類であった文書データを含めて処理を行うことで、本発明の構成を備えていない場合と比べて、文書分類の精度

が向上する。

[0029] 請求項 10 に記載の発明によると、本発明の構成を備えていない場合と比べて、文書分類の精度が向上する。

[0030] 請求項 11 に記載の発明によると、本発明の構成を備えていない場合と比べて、文書分類の精度が向上する。

図面の簡単な説明

[0031] [図1]本発明の実施形態に係る文書分類装置の一例を示すブロック図である。

[図2]動作例 1 を示すフローチャートである。

[図3]動作例 2 を示すフローチャートである。

[図4]予め特定のカテゴリに分類された学習用データの一例を示す表である。

[図5]学習用データのベクトル表現の一例を示す表である。

[図6]クラスタリングの結果を示す表である。

[図7]クラスタの分析結果を示す表である。

[図8]クラスタカテゴリが付与された学習用データの一例を示す表である。

[図9]変形例に係る文書分類装置の一例を示すブロック図である。

[図10]動作例 3 を示すフローチャートである。

[図11]動作例 4 を示すフローチャートである。

発明を実施するための形態

[0032] 図 1 に、本発明の実施形態に係る文書分類装置の一例を示す。文書分類装置 1 は、記憶部 10、特徴抽出部 16、クラスタリング部 18、クラスタ分析部 20、クラスタカテゴリ記憶部 22 及びカテゴリ更新部 24 を備え、予め特定のカテゴリに分類された学習用データに、カテゴリの定義の曖昧性を考慮した新たなカテゴリを付与する。

[0033] 記憶部 10 は、データ記憶部 12 とカテゴリ記憶部 14 とを含む。データ記憶部 12 には、予め特定のカテゴリに分類された学習用データが記憶されている。学習用データは、例えば、単一又は複数の文で構成された文書データである。学習用データは、予め特定のカテゴリに分類されていれば、カテゴリの種類や学習データの質にかかわらず、どのようなデータであってもよ

い。例えば、文書の章や段落のタイトルをカテゴリとし、章や段落内部の文書を学習用データとし、文書の構造を利用して自動で収集されるデータであってもよい。学習用データには、当該学習用データを識別するための学習用データ識別情報（例えばID番号等）が付与されており、また、当該学習用データのカテゴリを示すカテゴリ情報が関連付けられている。カテゴリ記憶部14は、学習用データに関連付けられたカテゴリ情報を記憶する。

[0034] 特徴抽出部16は、複数の学習用データのそれぞれから特徴情報を抽出し、抽出した特徴情報を用いて各学習用データをベクトル化する。特徴抽出部16は、形態素情報、品詞情報、構文・意味情報、文字・単語N-gram、任意の単語、句読点の数、及び、過去形の単語の数等のうちの1又は複数の特徴情報として抽出する。これらの情報はベクトルの素性となる。各素性のスコアは二値（0又は1）でもよいし、出現数やTf-idf（Term Frequency-Inverse Document Frequency）値等でもよい。また、特徴抽出部16は、辞書等を利用することで同義語を同一素性として扱ってもよいし、Tf-idf値に閾値を設けてノイズとなり得る素性を除去し、ベクトルの次元を圧縮してもよい。ノイズとなり得る素性を除去することで、文書分類の精度が向上する。例えば、idf値が下限閾値以下となる単語は一般的な単語であると判断され得るため、ベクトルの素性として使用しなくてもよい。また、idf値が上限閾値以上となる単語は、素性に識別器が大きく依存する可能性があるため、素性として使用しなくてもよい。以上のようにして、特徴抽出部16は、素性のスコアに応じて、クラスタリング部18によるクラスタ分けに用いられる特徴情報を選択してもよい。

[0035] クラスタリング部18は、特徴抽出部16によって抽出された特徴情報に基づき、類似した特徴情報を有する学習用データを同一のクラスタに分類する。具体的には、クラスタリング部18は、Repeated Bisect ion法、K-means法又はLDA（Latent Dirichlet Allocation）法等の手法を用いることで、ベクトル表現さ

れた学習用データをクラスタ分けする。例えば、クラスタリング部 18 は、特徴情報の出現頻度が類似する学習用データを同一のクラスタに分類する。クラスタリング部 18 は、クラスタを識別するためのクラスタ識別情報（例えば ID 番号等）を各クラスタに付与する。

[0036] クラスタ分析部 20 は、クラスタに属する学習用データに関連付けられたカテゴリ情報に基づき、当該クラスタ内のカテゴリの構成を分析し、分析結果をクラスタカテゴリ記憶部 22 に出力する。具体的には、クラスタ分析部 20 は、クラスタに属する学習用データに関連付けられたカテゴリ情報に基づき、同一クラスタに属する各学習用データのカテゴリの当該クラスタ内での出現数と、当該クラスタ内における出現割合とを求める。各カテゴリの出現割合は、同一クラスタ内に属する全カテゴリの出現数に対する各カテゴリの出現数である。また、クラスタを構成するデータ数が、全体又は他のクラスタのデータ数よりも少なく、その数の差が予め決定された閾値以上となるクラスタはノイズとなり得るため、クラスタ分析部 20 は、そのクラスタに属する学習用データを除去してもよい。また、同一クラスタに属するカテゴリのうち、数が他のカテゴリよりも少なく、その数の差が予め決定された閾値以上となるカテゴリはノイズとなり得るため、クラスタ分析部 20 は、そのカテゴリを除去してもよい。ノイズとなり得るクラスタや学習用データを除去することで、文書分類の精度が向上する。例えば、出現割合が予め決定された閾値以下となるカテゴリを、クラスタから除去してもよい。また、クラスタ分析部 20 は、1 つのクラスタに含まれるカテゴリの数を制限してもよい。例えば、クラスタ分析部 20 は、出現割合が大きいカテゴリから予め決定された数のカテゴリをクラスタに含ませ、それ以外のカテゴリをクラスタに含ませなくてもよい。以上のようにして、クラスタ分析部 20 は、カテゴリの出現割合に応じて、クラスタカテゴリに用いるカテゴリを選択してもよい。

[0037] クラスタカテゴリ記憶部 22 は、クラスタ分析部 20 の分析結果としてのクラスタカテゴリ情報を記憶する。

[0038] カテゴリ更新部 24 は、同一クラスタに属する各カテゴリと各カテゴリの出現割合とをクラスタカテゴリとして新たに定義し、このクラスタカテゴリを示すクラスタカテゴリ情報を用いることで、各学習用データのカテゴリを更新する。すなわち、カテゴリ更新部 24 は、同一クラスタに属する学習用データに、当該同一クラスタに属する別の学習用データのカテゴリも自己のカテゴリとして付与する。例えば、カテゴリ更新部 24 は、同一クラスタに属する各学習用データに、当該クラスタのクラスタ識別情報を付与し、当該クラスタのクラスタカテゴリ情報を関連付ける。このようにして、各学習用データをクラスタカテゴリに分類する。なお、各クラスタカテゴリには、クラスタ識別情報が付与される。カテゴリ記憶部 14 は、学習用データに関連付けられたクラスタカテゴリ情報を記憶する。このように、各学習用データにクラスタカテゴリが付与されることになるため、各学習用データは、元々のカテゴリに分類されているとともに、クラスタカテゴリに分類されてもいる。すなわち、各学習用データには、元々のカテゴリを示すカテゴリ情報とクラスタカテゴリを示すクラスタカテゴリ情報とが関連付けられていることになる。

[0039] なお、記憶部 10 及びクラスタカテゴリ記憶部 22 を文書分類装置 1 に設けずに、文書分類装置 1 の外部に設けてもよい。

[0040] (動作例 1)

次に、図 2 に示すフローチャートを参照して、文書分類装置 1 の動作例 1 を説明する。まず、特徴抽出部 16 は、複数の学習用データをデータ記憶部 12 から読み出し、複数の学習用データを解析することで複数の学習用データのそれぞれから特徴情報を抽出し、各学習用データをベクトル化する (S01)。次に、クラスタリング部 18 は、ベクトル表現された複数の学習用データをクラスタに分類する (S02)。そして、クラスタ分析部 20 は、クラスタに属する学習用データに関連付けられたカテゴリ情報に基づき、当該クラスタ内のカテゴリの構成を分析する (S03)。カテゴリ更新部 24 は、同一クラスタに属する各カテゴリと各カテゴリの出現割合とをクラス

タカテゴリとして定義し、同一クラスタに属する各学習用データに当該クラスタのクラスタカテゴリ情報を関連付けることで、各学習用データのカテゴリを更新する（S 0 4）。

[0041] （動作例 2）

次に、図 3 に示すフローチャートを参照して、文書分類装置 1 の動作例 2 を説明する。まず、特徴抽出部 1 6 は、複数の学習用データを解析することで複数の学習用データのそれぞれから特徴情報を抽出する（S 1 0）。次に、特徴抽出部 1 6 は、抽出した特徴情報のうちノイズとなり得る特徴情報を除去し、ノイズとなり得ない特徴情報を選択して各学習用データをベクトル化する（S 1 1）。例えば、特徴抽出部 1 6 は、抽出した特徴情報のうち同義語を同一素性として扱い、又は、T f - i d f 値に基づいて不要な素性を除去し、各学習用データをベクトル化する。そして、クラスタリング部 1 8 は、ベクトル表現された複数の学習用データをクラスタに分類し（S 1 2）、クラスタ分析部 2 0 は、クラスタ内のカテゴリの構成を分析する（S 1 3）。また、クラスタ分析部 2 0 は、ノイズとなり得るクラスタの学習データを除去する（S 1 4）。クラスタ分析部 2 0 は、数が他のカテゴリよりも少なく、その数の差が閾値以上となるカテゴリを除去してもよい。そして、カテゴリ更新部 2 4 は、同一クラスタに属する各学習用データにクラスタカテゴリ情報を関連付けることで、各学習用データのカテゴリを更新する（S 1 5）。

[0042] （具体例）

次に、文書分類装置 1 の動作について具体例を挙げて説明する。図 4 に、予め特定のカテゴリに分類された学習用データの一例を示す。例えば、学習用データ識別情報（I D = 1, 2）が付された学習用データは「概要」というカテゴリに予め分類されており、I D = 3 の学習用データは「経歴」というカテゴリに予め分類されており、I D = 4 の学習用データは「構成」というカテゴリに予め分類されている。各学習用データは記憶部 1 0 に記憶されており、各学習用データのカテゴリを示すカテゴリ情報は、カテゴリ記憶部

14に記憶されている。

[0043] 特徴抽出部16は、各学習用データ（ID=1, 2, 3, ...）から特徴情報を抽出し、各学習用データをベクトル化する。図5に、ID=1の学習用データのベクトル表現の一例を示す。説明を簡略にするために、内容語（単語）をベクトルの素性として用い、出現数をスコアとして用いている。図5中のWIDは、特徴抽出部16によって各単語に付けられた識別情報である。なお、WIDはベクトルのインデックス番号を表し、出現数はインデックス番号におけるスコアを表す。特徴抽出部16は、辞書等を利用することで、同義語に同一のWIDを付与してもよい。また、スコア（出現数、出現頻度）に対して閾値を設定し、スコアが閾値以下となる素性を除去してもよい。すなわち、スコアに応じて、クラスタリング部18によるクラスタ分けに用いられる特徴情報を選択してもよい。

[0044] クラスタリング部18は、ベクトル表現された学習用データを、予め決定された手法によってクラスタに分類する。図6にクラスタリングの結果を示す。図6中のCIDは、クラスタを識別するためのクラスタ識別情報である。図6に示す例では、CID=1のクラスタには、ID=1, 3, 10, ...の学習用データが分類されている。ID=1の学習用データは予め「概要」カテゴリに分類され、ID=3の学習用データは予め「経歴」カテゴリに分類されているため、「概要」カテゴリ及び「経歴」カテゴリがCID=1のクラスタに属することになる。また、CID=2のクラスタには、ID=2の学習用データが分類されている。ID=2の学習用データは予め「概要」カテゴリに分類されているため、「概要」カテゴリがCID=2のクラスタに属することになる。他のCIDのクラスタについても同様に、各クラスタに分類された学習用データのカテゴリが、対応するクラスタに属することになる。以下では、説明の便宜上、CID=1のクラスタを「クラスタ1」と称し、CID=2のクラスタを「クラスタ2」と称する。他のCIDのクラスタについても同様である。

[0045] クラスタ分析部20は、クラスタ1に属する各学習用データ（ID=1,

3, 10, ...) のカテゴリのクラスタ 1 内での出現数と、クラスタ 1 内における各カテゴリの出現割合とを求める。他のクラスタ 2, 3, ... についても同様に、カテゴリの出現数と出現割合とを求める。クラスタ内のカテゴリの出現数は、当該カテゴリに予め分類され、クラスタリング部 18 によって当該クラスタに分類された学習用データの数意味する。なお、クラスタ分析部 20 は、ノイズとなり得るクラスタや学習用データを除去してもよい。図 7 にクラスタの分析結果を示す。図 7 に示す例では、クラスタ 1 には 516 個の学習用データ（カテゴリ）が分類されている（要素数 = 516）。具体的には、クラスタ 1 には、120 個の「経歴」カテゴリが含まれ、101 個の「概要」カテゴリが含まれている。これは、「経歴」カテゴリに予め分類された 120 個の学習用データが、クラスタリング部 18 によってクラスタ 1 に分類され、「概要」カテゴリに予め分類された 101 個の学習用データが、クラスタリング部 18 によってクラスタ 1 に分類されたことを意味する。また、クラスタ 1 においては、「経歴」カテゴリの出現割合は 0.23（23%）であり、「概要」カテゴリの出現割合は 0.20（20%）である。すなわち、クラスタ 1 に分類された全カテゴリのうち 23% のカテゴリが「経歴」カテゴリであり、20% のカテゴリが「概要」カテゴリであることを意味する。

[0046] クラスタ 2 には 1 個の学習用データ（カテゴリ）が分類されている（要素数 = 1）。具体的には、1 個の「概要」カテゴリがクラスタ 2 に分類されている。これは、「概要」カテゴリに予め分類された 1 個の学習用データが、クラスタリング部 18 によってクラスタ 2 に分類されたことを意味する。従って、クラスタ 2 においては、「概要」カテゴリの出現割合は 1.00（100%）である。

[0047] クラスタ 3 には 333 個の学習用データ（カテゴリ）が分類されている（要素数 = 333）。具体的には、クラスタ 3 には、91 個の「構成」カテゴリが含まれ、52 個の「概要」カテゴリが含まれている。これは、「構成」カテゴリに予め分類された 91 個の学習用データが、クラスタリング部 18

によってクラスタ 3 に分類され、「概要」カテゴリに予め分類された 5 2 個の学習用データが、クラスタリング部 1 8 によってカテゴリ 3 に分類されたことを意味する。また、クラスタ 3 においては、「構成」カテゴリの出現割合は 0. 2 7 (2 7 %) であり、「概要」カテゴリの出現割合は 0. 1 6 (1 6 %) である。

[0048] カテゴリ更新部 2 4 は、クラスタ 1 に属する各カテゴリ（経歴、概要、・・・）と各カテゴリの出現割合（0. 2 3、0. 2 0、・・・）とを、C I D = 1 のクラスタカテゴリとして新たに定義する。同様に、カテゴリ更新部 2 4 は、クラスタ 2 に属する「概要」カテゴリと出現割合（1. 0 0）とを、C I D = 2 のクラスタカテゴリとして新たに定義し、クラスタ 3 に属する各カテゴリ（構成、概要、・・・）と各カテゴリの出現割合（0. 2 7、0. 1 6、・・・）とを、C I D = 3 のクラスタカテゴリとして新たに定義する。他のクラスタについても同様である。

[0049] カテゴリ更新部 2 4 は、クラスタ 1 に属する学習用データに、クラスタ 1 のクラスタ識別情報（C I D = 1）を付与し、C I D = 1 のクラスタカテゴリを示すクラスタカテゴリ情報を関連付ける。他のクラスタに属する学習用データについても同様に、クラスタカテゴリ情報を関連付ける。以下では、説明の便宜上、C I D = 1 のクラスタカテゴリを「クラスタカテゴリ 1」と称し、C I D = 2 のクラスタカテゴリを「クラスタカテゴリ 2」と称する。他の C I D のクラスタカテゴリについても同様である。

[0050] 図 8 に、クラスタカテゴリが付与された学習用データの一例を示す。I D = 1, 3 の学習用データはクラスタ 1 に属しているため、クラスタカテゴリ 1 を示すクラスタカテゴリ情報が関連付けられている。すなわち、「経歴」カテゴリ、「経歴」カテゴリの出現割合、「概要」カテゴリ及び「概要」カテゴリの出現割合を含むクラスタカテゴリ情報が、I D = 1, 3 の学習用データに関連付けられている。また、I D = 2 の学習用データはクラスタ 2 に属しているため、クラスタカテゴリ 2 を示すクラスタカテゴリ情報が関連付けられている。すなわち、「概要」カテゴリ及び「概要」カテゴリの出現割

合を示すクラスカテゴリ情報が、ID = 2 の学習用データに関連付けられている。

[0051] ID = 1 の学習用データを例にして説明すると、ID = 1 の学習用データに付与されたクラスカテゴリは、例えば、以下のように解釈され得る。

(1) ID = 1 の学習用データは、23%の確率で「経歴」カテゴリ、20%の確率で「概要」カテゴリに分類されるデータである。

(2) ID = 1 の学習用データは、23%の「経歴」カテゴリの特徴と20%の「概要」カテゴリの特徴とを有するデータである。

(3) ID = 1 の学習用データは、「経歴」カテゴリの特徴と「概要」カテゴリの特徴とを有するデータである。

ID = 1 の学習用データは「概要」カテゴリに予め分類されていたが、「経歴」カテゴリの出現割合（23%）が「概要」カテゴリの出現割合（20%）よりも大きいため、「概要」カテゴリよりも「経歴」カテゴリの特徴が強いと分析されている。他のクラスカテゴリについても同様に解釈され得る。

[0052] 以上のように、予め特定のカテゴリに分類された学習用データの特徴情報に基づいて学習用データをクラスタに分類し、クラスタに属するカテゴリと当該カテゴリの出現割合とをクラスタカテゴリとして新たに定義して学習用データに関連付けることで、カテゴリ間の境界が曖昧でカテゴリの定義に曖昧性がある場合であっても、その曖昧性を表現したクラスタカテゴリが学習用データに付与されることになる。言葉は様々な意味や概念を持つため、各カテゴリが互いに独立していないことがあり、学習用データに予め付与されたカテゴリは、当該学習用データの特徴を正確に表していないことがある。本実施形態によると、カテゴリの曖昧性を表現したクラスタカテゴリが学習用データに付与されるため、当該学習用データの特徴をより正確に表したカテゴリが学習用データに付与されることになる。

[0053] 例えば、上述したID = 1 の学習用データのように、元々は「概要」カテゴリが付与されていたが、「経歴」の要素が「概要」よりも強い可能性があ

る。この場合に、「概要」カテゴリ又は「経歴」カテゴリのいずれか一方のみを $ID = 1$ の学習用データに付与するのではなく、出現割合も含めて両カテゴリをクラスタカテゴリとして $ID = 1$ の学習用データに付与することで、クラスタカテゴリを付与しない場合と比べて、 $ID = 1$ の学習用データの特徴がより正確に表現されることになる。例えば、「概念」カテゴリだけが $ID = 1$ の学習用データに付与されている場合、 $ID = 1$ の学習用データが「経歴」カテゴリの特徴を有しているにもかかわらず、その特徴が表されない。これに対して、クラスタカテゴリを学習用データに付与することで、 $ID = 1$ の学習用データが「概要」及び「経歴」カテゴリの特徴を有していることが表されるため、 $ID = 1$ の学習用データの特徴がより正確に表現されることになる。

[0054] 以上のように、予め特定のカテゴリに分類された学習用データを対象にして分類処理を行い、学習用データを新たなクラスタカテゴリに分類することで、学習用データを用いた文書分類の精度が向上し得る。

[0055] なお、図示しない表示装置によって、カテゴリ、クラスタカテゴリ及び学習用データの対応関係を表示してもよい。例えば図8に示す表形式で、その対応関係を表示してもよい。これにより、ユーザが学習用データの概念や意味を把握するための情報や、ユーザがカテゴリを付与し直すための情報が、ユーザに提供されることになる。

[0056] (変形例)

次に、図9を参照して、変形例に係る文書分類装置について説明する。変形例に係る文書分類装置1Aは、上記の文書分類装置1の構成に加えて、処理部30、入力部42及び出力部44を備えている。処理部30は、識別器作成部32、カテゴリ分類部34、検索部36及び学習用データ更新部38を備えている。未分類データ記憶部46には、特定のカテゴリに分類されていない文書データ（「未分類データ」と称する）が記憶されている。

[0057] 入力部42は例えばユーザインターフェースであり、ユーザによる情報の入力を受け付ける。出力部44は、例えば表示装置、ネットワークインター

フェース又は印刷装置等であり、処理部 30 の処理結果を出力する。

[0058] 識別器作成部 32 は、特定のカテゴリに分類された学習用データから識別器（「カテゴリ識別器」）を作成する。例えば、学習用データの主成分ベクトルを識別器として用いる。また、識別器作成部 32 は、クラスタカテゴリに分類された学習用データから識別器（「クラスタカテゴリ識別器」）を作成してもよい。

[0059] カテゴリ分類部 34 は、識別器を用いて未分類データのカテゴリを求め、未分類データをカテゴリに分類する。また、カテゴリ分類部 34 は、識別器を用いて未分類データのクラスタカテゴリを求め、未分類データをクラスタカテゴリに分類してもよい。

[0060] 検索部 36 は、クラスタカテゴリが付与された学習用データを検索対象とし、入力部 42 が受け付けた検索クエリを用いてデータを検索する。検索部 36 は、学習用データと検索クエリとの一致度によってスコアを算出し、スコア降順で検索結果（より高いスコアのデータをより上位とする結果）を出力する。さらに、検索クエリにカテゴリの用語が含まれている場合、検索部 36 は、クラスタカテゴリも検索対象とし、クラスタカテゴリに属するカテゴリの構成に応じて、検索クエリと検索対象データとの一致度を表すスコアを調整する。例えば、検索クエリに含まれる用語と一致するカテゴリの出現割合に応じた重みをスコアに付けることで、スコアを調整する。これにより、検索クエリに含まれる用語と一致するカテゴリの出現割合が大きいほど、当該カテゴリに分類された学習用データのスコアは高くなる。

[0061] 学習用データ更新部 38 は、未分類データにカテゴリ又はクラスタカテゴリが付与された場合、各種データを更新する。例えば、学習用データ更新部 38 は、カテゴリやクラスタカテゴリが付与された未分類データを学習用データとしてデータ記憶部 12 に記憶させる。また、学習用データ更新部 38 は、カテゴリ記憶部 14 やクラスタカテゴリ記憶部 22 に記憶されているデータを更新してもよい。また、学習用データ更新部 38 は、検索部 36 の検索結果に基づいて学習用データのカテゴリを更新してもよい。

[0062] (動作例3)

次に、図10を参照して、変形例に係る文書分類装置1Aの動作例(動作例3)を説明する。動作例3では、未分類データにカテゴリやクラスタカテゴリを付与して学習用データを更新する。

[0063] まず、カテゴリ分類部34は、識別器作成部32によって作成されたクラスタカテゴリ識別器を用いて未分類データのクラスタカテゴリを求め、未分類データをクラスタカテゴリに分類する(S20)。そして、クラスタカテゴリに属するカテゴリからカテゴリを選択する(S21)。例えば、カテゴリ分類部34は、クラスタカテゴリに属するカテゴリのうち、出現割合が最大のカテゴリを、未分類データに付与してもよい。出現割合が最大のカテゴリは、当該未分類データの特徴をより正確に表している可能性があるため、出現割合が最大のカテゴリを未分類データに付与することで、未分類データの特徴をより正確に表すカテゴリが未分類データに付与されることになる。または、出力部44がクラスタカテゴリに属するカテゴリの一覧を表示し、ユーザが入力部42を用いてカテゴリを選択してもよい。この場合、カテゴリ分類部34は、ユーザによって選択されたカテゴリを未分類データに付与する。なお、クラスタカテゴリのみを未分類データに付与してもよい。

[0064] そして、学習用データ更新部38は、各種データを更新する(S22)。例えば、学習用データ更新部38は、カテゴリやクラスタカテゴリが付与された未分類データを学習用データとしてデータ記憶部12に記憶させる。また、学習用データ更新部38は、選択されたカテゴリの出現数を1つインクリメントし、クラスタの要素数及びカテゴリの出現割合を更新する。これにより、カテゴリ記憶部14及びクラスタカテゴリ記憶部22に記憶されているデータが更新される。図7を参照して説明すると、ステップS21にて、クラスタカテゴリ1(CID=1)に属する「経歴」カテゴリが選択された場合、学習用データ更新部38は、クラスタカテゴリ1における「経歴」カテゴリの出現数を1つインクリメントし、クラスタカテゴリ1の要素数を1つインクリメントし、クラスタカテゴリ1における「経歴」カテゴリの出現

割合を更新する。なお、未分類データにカテゴリが付与された場合、識別器作成部 32 は識別器を更新してもよい。

[0065] 以上のように、未分類のデータにカテゴリやクラスタカテゴリを付与して学習用データを更新することで、クラスタカテゴリの作成の基礎となる情報が更新されるため、クラスタカテゴリの分類の精度が向上する。

[0066] また、別の例として、カテゴリや段落の見出し等が付与されていない文書データに、カテゴリや見出しを付与したい場合、動作例 3 を適用してもよい。例えば、カテゴリ分類部 34 は、識別器を用いて対象データをクラスタカテゴリに分類し、クラスタカテゴリ、クラスタカテゴリに属するカテゴリのうち出現割合が最大のカテゴリ、又は、クラスタカテゴリに属するカテゴリのうちユーザによって選択されたカテゴリを、対象データに付与する。この場合も、学習用データ更新部 38 は、各種データを更新する。

[0067] なお、動作例 3 によって文書分類装置 1A が動作する場合、処理部 30 は検索部 36 を備えていなくてもよい。

[0068] (動作例 4)

次に、図 11 を参照して、変形例に係る文書分類装置 1A の別の動作例 (動作例 4) を説明する。動作例 4 では、クラスタカテゴリが付与された学習用データを検索対象として検索を行い、その検索結果に基づいて学習用データを更新する。

[0069] まず、入力部 42 を介して検索クエリが入力されると、検索部 36 は、クラスタカテゴリが付与された学習用データを検索対象とし、検索対象の学習用データと検索クエリとの一致度によってスコアを算出する (S30)。

[0070] 検索クエリにカテゴリの用語 (例えば、「概要」や「経歴」等) が含まれていない場合 (S31, No)、検索部 36 は、ステップ S30 での検索結果を出力し、検索処理は終了する。例えば、出力部 44 はスコア降順で検索結果を表示する。

[0071] 一方、検索クエリにカテゴリの用語が含まれている場合 (S31, Yes)、検索部 36 は、クラスタカテゴリも検索対象として検索する (S32)

。図8を参照して説明すると、例えば検索クエリが「経歴」というカテゴリの用語を含む場合、検索部36は、クラスタカテゴリを検索対象として検索を行い、検索クエリの用語「経歴」を含むクラスタカテゴリを特定し、当該クラスタカテゴリが付与されている学習用データを抽出する。図8に示す例では、クラスタカテゴリ1が付与されているID=1, 3の学習用データが抽出される。仮にID=1, 3の学習用データのテキストに「経歴」という用語が含まれていなくても、ID=1, 3の学習用データに付与されたクラスタカテゴリ1には「経歴」カテゴリが含まれているため、ID=1, 3の学習用データが検索によって抽出されることになる。

[0072] そして、検索部36は、検索クエリの用語と一致するカテゴリの出現割合が大きいほど、当該カテゴリに分類された学習用データのスコアを高くし、ステップS30での検索結果を再ランキングする(S33)。例えば、出力部44は、再ランキングされた検索結果を表示する。

[0073] ユーザは再ランキングされた検索結果を参照し、入力部42を用いて任意の学習用データを選択する(S34)。

[0074] ユーザによって学習用データが選択されると、学習用データ更新部38は各種データを更新する(S35)。この更新処理について、図8を参照して説明する。例えば、検索クエリに「経歴」というカテゴリの用語が含まれ、検索結果の中から「概要」カテゴリに分類されたID=1の学習用データがユーザによって選択された場合、学習用データ更新部38は、ID=1の学習用データのカテゴリを「概要」カテゴリから「経歴」カテゴリに変更する。このようにカテゴリを変更することで、ID=1の学習用データのカテゴリが修正され、ID=1の学習用データの特徴をより正確に表すカテゴリがID=1の学習用データに付与されることになる。また、動作例3と同様に、学習用データ更新部38は、カテゴリの選択に応じて、クラスタの要素数、カテゴリの出現数及び出現割合を更新する。

[0075] 以上のようにクラスタカテゴリも検索対象とすることで、検索クエリが学習用データのテキストに存在せず、学習用データのみを検索対象とした場合

には抽出されることのない学習用データも抽出される。これにより、検索クエリと関係があり得る学習用データが抽出される。例えば、学習用データのテキストに検索クエリが含まれておらず、検索クエリが学習用データの概念や意味を示す場合、学習用データのみを検索対象とすると、検索クエリと概念や意味が一致する学習用データは抽出されないが、クラスタカテゴリも検索対象とすることで、検索クエリと概念や意味が一致する学習用データが抽出されることになる。従って、単に学習用データのテキストを対象として検索する場合と異なり、概念や意味をも含めて学習用データを検索しているともいえる。

[0076] 上記の文書分類装置 1, 1 A は、一例としてハードウェア資源とソフトウェアとの協働により実現される。具体的には、文書分類装置 1, 1 A は、図示しない CPU 等のプロセッサを備えている。プロセッサは、図示しない記憶装置に記憶されたプログラムを読み出して実行することにより、上述した特徴抽出部 16、クラスタリング部 18、クラスタ分析部 20、カテゴリ更新部 24 及び処理部 30 のそれぞれの機能を実行する。上記プログラムは、CD や DVD 等の記録媒体を経由して、又は、ネットワーク等の通信手段を経由して、ハードディスクドライブ (HDD) 等の記憶装置に記憶される。なお、上記プログラムは、ハードディスクドライブ等の記憶装置に予め記憶されていてもよい。ハードディスクドライブ等の記憶装置に記憶されたプログラムが、RAM 等のメモリに読み出されてプロセッサによって実行されることにより、上述した各部の機能が実現される。

符号の説明

[0077] 1, 1 A 文書分類装置、10 記憶部、12 データ記憶部、14 カテゴリ記憶部、16 特徴抽出部、18 クラスタリング部、20 クラスタ分析部、22 クラスタカテゴリ記憶部、24 カテゴリ更新部、30 処理部、32 識別器作成部、34 カテゴリ分類部、36 検索部、38 学習用データ更新部、42 入力部、44 出力部、46 未分類データ記憶部。

請求の範囲

- [請求項1] 予め特定のカテゴリに分類された複数の文書データのそれぞれから特徴情報を抽出する特徴抽出手段と、
- 前記特徴情報の出現頻度が類似する文書データを同一クラスタに分類するクラスタリング手段と、
- 同一クラスタに分類された文書データに、前記同一クラスタに分類された他の文書データのカテゴリも自己のカテゴリとして付与するカテゴリ更新手段と、
- を有することを特徴とする文書分類装置。
- [請求項2] 請求項1に記載の文書分類装置であって、
- 前記特徴抽出手段は、前記特定のカテゴリに関連する情報を特徴情報として前記複数の文書データのそれぞれから抽出する、
- ことを特徴とする文書分類装置。
- [請求項3] 請求項1又は請求項2に記載の文書分類装置であって、
- 前記カテゴリ更新手段は、前記同一クラスタにおける各文書データのカテゴリの出現割合を、前記同一クラスタに分類された文書データに付与する、
- ことを特徴とする文書分類装置。
- [請求項4] 請求項1から請求項3のいずれか一項に記載の文書分類装置であって、
- 検索クエリを受け、前記検索クエリにカテゴリに関する情報が含まれている場合、文書データに予め付与されたカテゴリ及び前記カテゴリ更新手段によって付与されたカテゴリも検索対象とし、前記検索クエリを用いて検索し、検索結果を出力する検索手段を更に有する、
- ことを特徴とする文書分類装置。
- [請求項5] 請求項4に記載の文書分類装置であって、
- 前記検索手段は、前記同一クラスタにおける各文書データのカテゴリの出現割合に応じた検索結果を出力する、

ことを特徴とする文書分類装置。

[請求項6]

請求項4又は請求項5に記載の文書分類装置であって、

前記検索クエリによって検索された文書データのカテゴリと前記検索クエリに含まれるカテゴリに関する情報とが一致しない場合、前記検索された文書データのカテゴリを、前記検索クエリに含まれるカテゴリに変更する更新手段を更に有する、

ことを特徴とする文書分類装置。

[請求項7]

請求項1から請求項6のいずれか一項に記載の文書分類装置であって、

前記カテゴリ更新手段によってカテゴリが付与された文書データに基づいて作成された識別器を用いて未分類の文書データにカテゴリを付与するカテゴリ分類手段を更に有する、

ことを特徴とする文書分類装置。

[請求項8]

請求項7に記載の文書分類装置であって、

前記カテゴリ分類手段は、出現割合が最大のカテゴリを前記未分類の文書データに付与する、

ことを特徴とする文書分類装置。

[請求項9]

請求項7又は請求項8に記載の文書分類装置であって、

前記カテゴリ分類手段によって処理された前記未分類の文書データを、前記特徴抽出手段、前記クラスタリング手段及び前記カテゴリ更新手段の処理対象の文書データに追加する手段を更に有する、

ことを特徴とする文書分類装置。

[請求項10]

請求項1から請求項9のいずれか一項に記載の文書分類装置であって、

前記特徴抽出手段は、文書データにおける特徴情報の出現頻度に応じて、前記クラスタリング手段によるクラスタ分けに用いられる特徴情報を選択する、

ことを特徴とする文書分類装置。

[請求項11] 請求項1から請求項10のいずれか一項に記載の文書分類装置であって、

前記カテゴリ更新手段は、前記同一カテゴリにおける各文書データのカテゴリの出現割合に応じてカテゴリを選択して文書データに付与する、

ことを特徴とする文書分類装置。

[請求項12] コンピュータに、

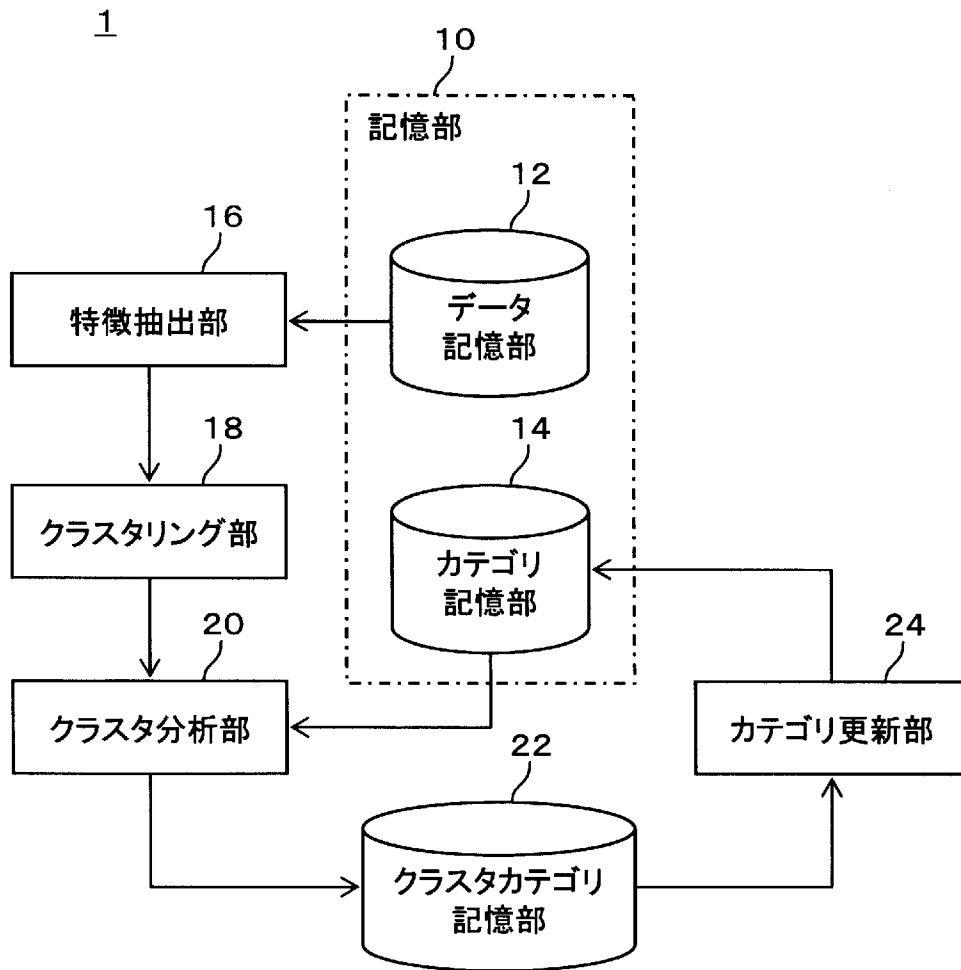
予め特定のカテゴリに分類された複数の文書データのそれぞれから特徴情報を抽出するステップと、

前記特徴情報の出現頻度が類似する文書データを同一クラスタに分類するステップと、

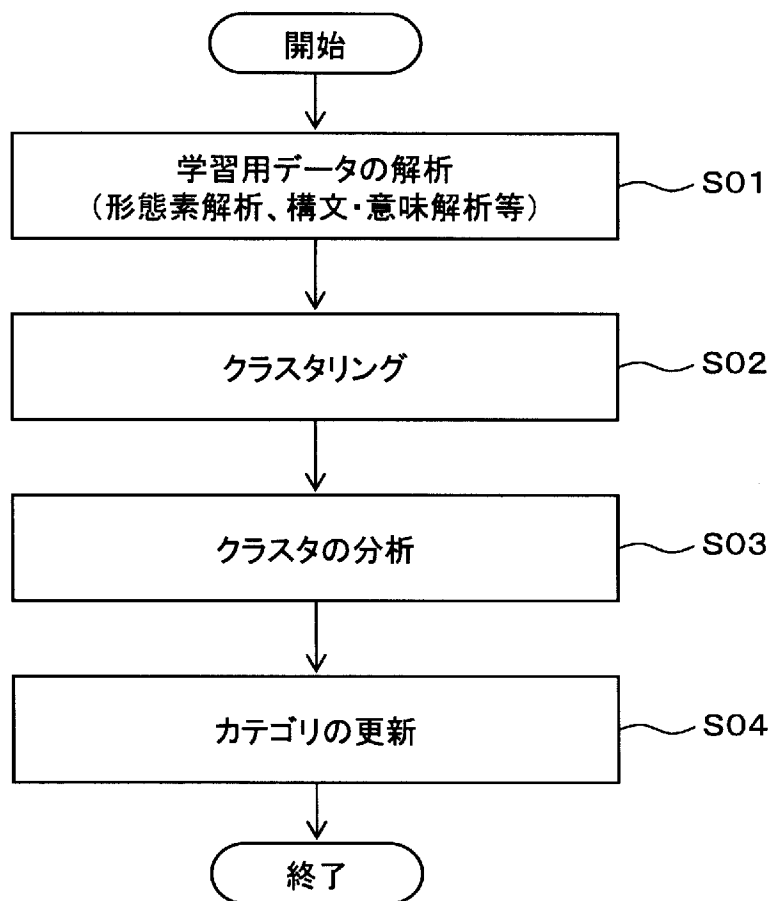
同一クラスタに分類された文書データに、前記同一クラスタに分類された他の文書データのカテゴリも自己のカテゴリとして付与するステップと、

を実行させることを特徴とするプログラム。

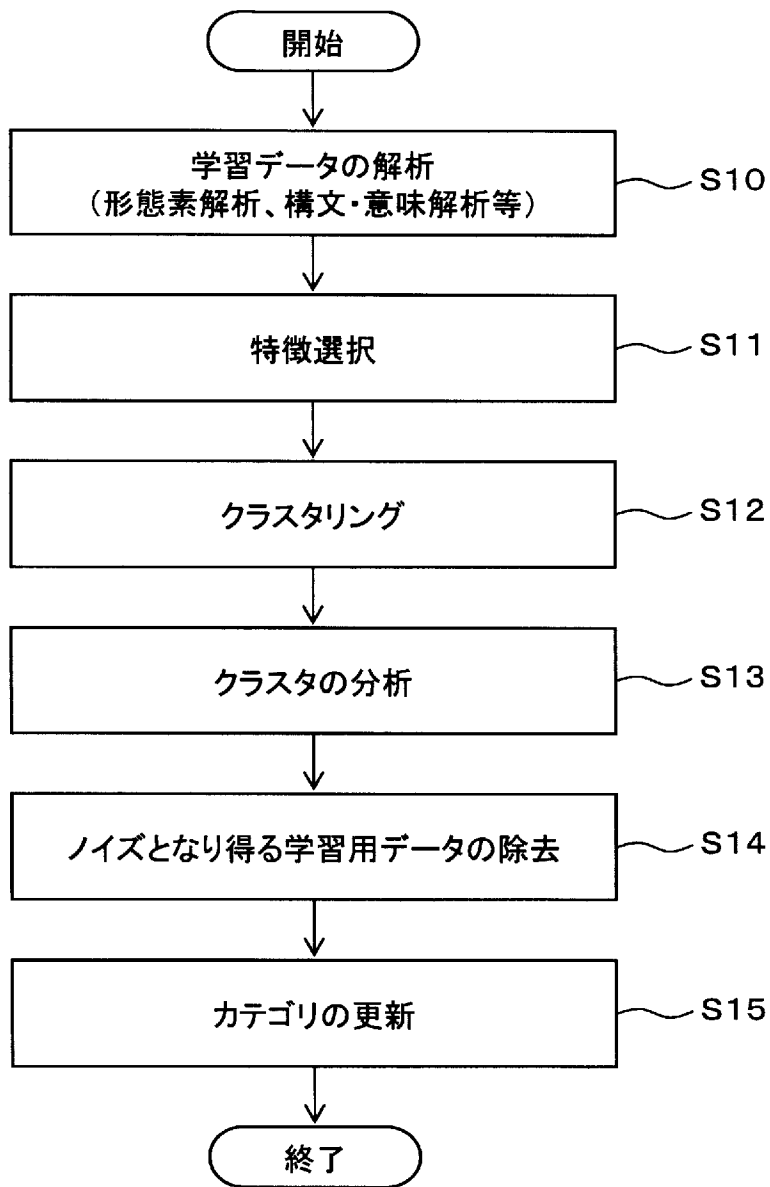
[図1]



[図2]



[図3]



[図4]

ID	カテゴリ	学習用データ
1	概要	
2	概要	
3	経歴	
4	構成	
...	...	...

[図5]

学習用データのベクトル化

WID	17	21	37	61	65	101	102	...
単語	FIA	1997年	発表する	ワールドラリーカー	生産モデル	内容	フォード	...
出現数	3	2	1	3	1	1	3	...

[図6]

クラスタリング結果

CID	分類された学習用データのID
1	1, 3, 10, ...
2	2
3	4, 5, 11, ...
...	

[図7]

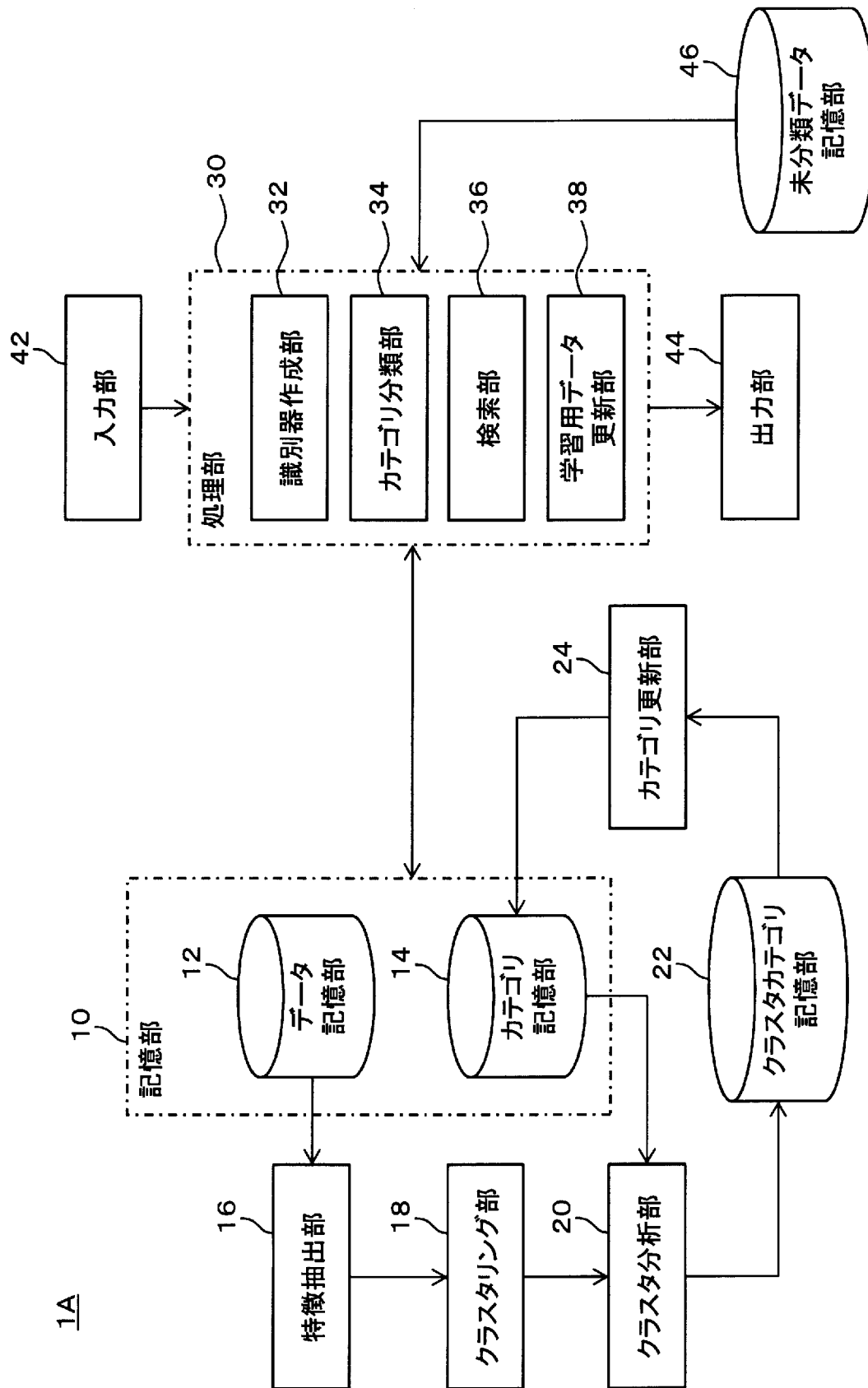
クラスタの分析

CID[要素数]	分類されたカテゴリ[出現数, 出現割合]
1[516]	経歴[120, 0. 23]、概要[101, 0. 20], ...
2[1]	概要[1, 1. 00]
3[333]	構成[91, 0. 27], 概要[52, 0. 16], ...
...	

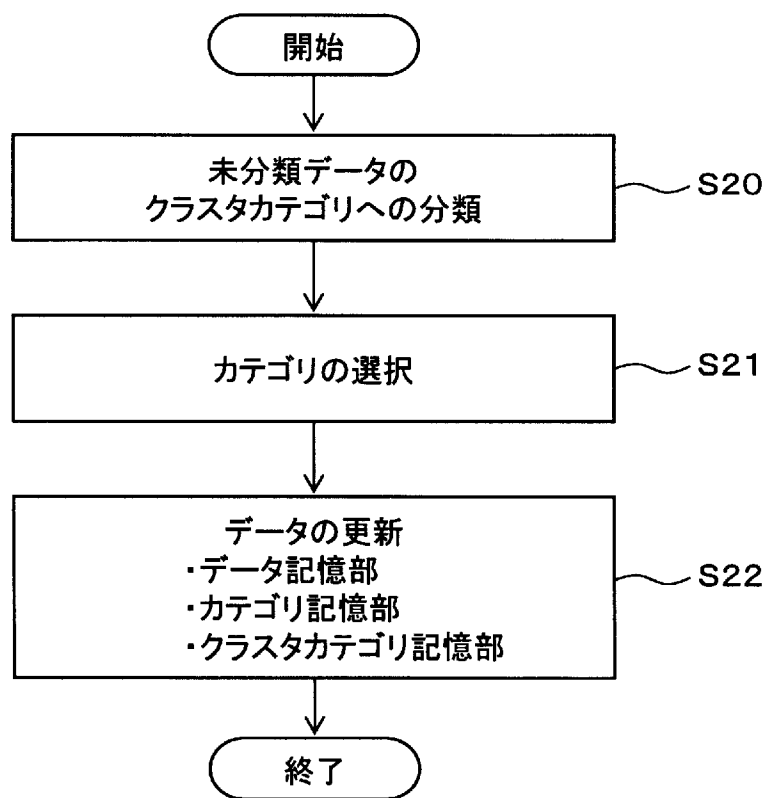
[図8]

ID	カテゴリ	クラスタカテゴリ	学習用データ
1	概要	1 経歴[0. 23], 概要[0. 20], ...	
2	概要	2 概要[1. 00]	
3	経歴	1 経歴[0. 23], 概要[0. 20], ...	
4	構成	3 構成[0. 27], 概要[0. 16], ...	
...	...		...

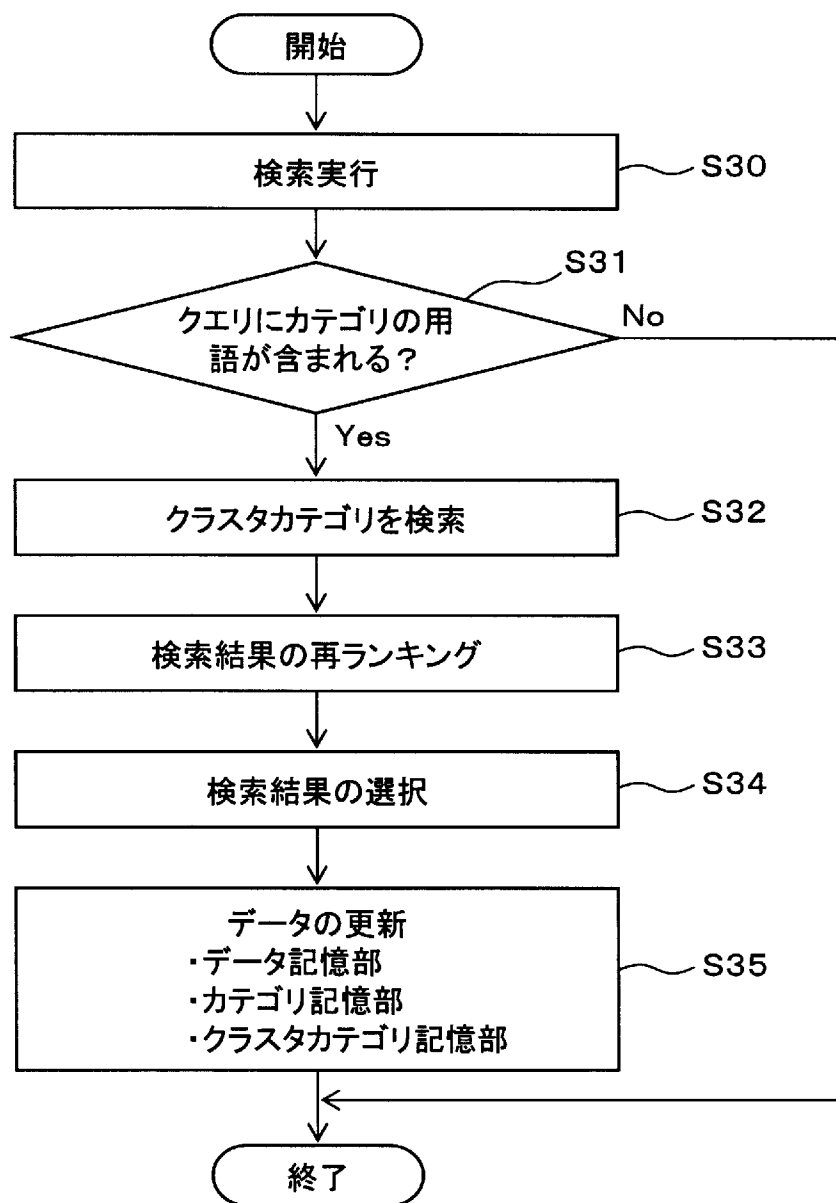
[図9]



[図10]



[図11]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/068852

A. CLASSIFICATION OF SUBJECT MATTER

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)

G06F17/30

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

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2013
Kokai Jitsuyo Shinan Koho	1971-2013	Toroku Jitsuyo Shinan Koho	1994-2013

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 A	JP 2009-259250 A (NEC (China) Co., Ltd.), 05 November 2009 (05.11.2009), entire text & CN 101561805 A	1-2, 7, 9-10, 12 3-6, 8, 11
X A	JP 2002-041544 A (Toshiba Corp.), 08 February 2002 (08.02.2002), entire text & US 2002/0016798 A1	1-2, 7, 9-10, 12 3-6, 8, 11
A	JP 10-078971 A (Canon Inc.), 24 March 1998 (24.03.1998), entire text (Family: none)	1-12

☒ 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
12 August, 2013 (12.08.13)

Date of mailing of the international search report
20 August, 2013 (20.08.13)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/068852

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 2009-070321 A (Fuji Xerox Co., Ltd.), 02 April 2009 (02.04.2009), entire text (Family: none)	1-12

A. 発明の属する分野の分類 (国際特許分類 (I P C))

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 (I P C))

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	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 3 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 3 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 3 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 A	JP 2009-259250 A (エヌイーシー (チャイナ) カンパニー, リミテッド) 2009.11.05 , 全文 & CN 101561805 A	1-2, 7, 9-10 , 12 3-6, 8, 11
X A	JP 2002-041544 A (株式会社東芝) 2002.02.08, 全文 & US 2002/0016798 A1	1-2, 7, 9-10 , 12 3-6, 8, 11

☒ 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

1 2 . 0 8 . 2 0 1 3

国際調査報告の発送日

2 0 . 0 8 . 2 0 1 3

国際調査機関の名称及びあて先

日本国特許庁 (I S A / J P)

郵便番号 1 0 0 - 8 9 1 5

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)

梅本 達雄

5 M

9 6 4 8

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 9 9

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 10-078971 A (キヤノン株式会社) 1998.03.24, 全文 (ファミリーなし)	1-12
A	JP 2009-070321 A (富士ゼロックス株式会社) 2009.04.02, 全文 (ファミリーなし)	1-12