

(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
2018年6月21日(21.06.2018)



(10) 国際公開番号

WO 2018/110029 A1

(51) 国際特許分類:
G06F 17/30 (2006.01) G06F 17/27 (2006.01)

(21) 国際出願番号: PCT/JP2017/034583

(22) 国際出願日: 2017年9月25日(25.09.2017)

(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 2016-240931 2016年12月13日(13.12.2016) JP

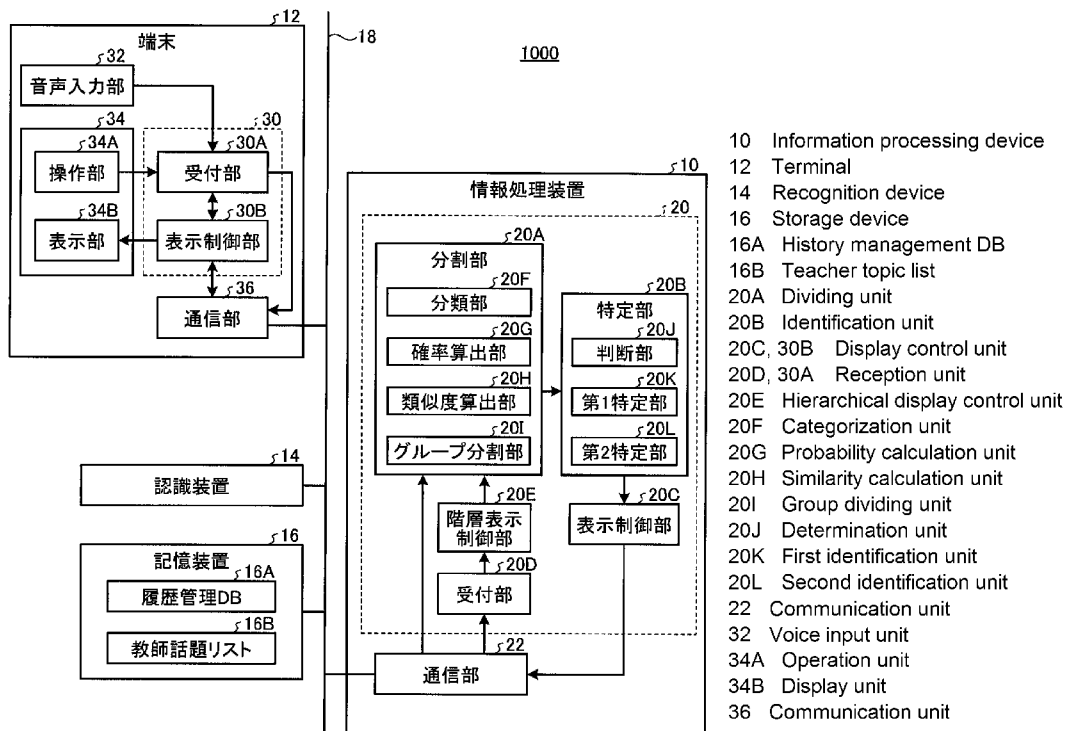
(71) 出願人: 株式会社東芝 (KABUSHIKI KAISHA TOSHIBA) [JP/JP]; 〒1058001 東京都港区芝浦一丁目1番1号 Tokyo (JP). 東芝デジタルソリューションズ株式会社 (TOSHIBA DIGITAL SOLUTIONS

CORPORATION) [JP/JP]; 〒2128585 神奈川県川崎市幸区堀川町7番地34 Kanagawa (JP).

(72) 発明者: 宮部 泰成 (MIYABE, Yasunari); 〒2128585 神奈川県川崎市幸区堀川町7番地34 東芝デジタルソリューションズ株式会社内 Kanagawa (JP). 後藤 和之 (GOTO, Kazuyuki); 〒2128585 神奈川県川崎市幸区堀川町7番地34 東芝デジタルソリューションズ株式会社内 Kanagawa (JP). 長 健太 (CHO, Kenta); 〒2128585 神奈川県川崎市幸区堀川町7番地34 東芝デジタルソリューションズ株式会社内 Kanagawa (JP). 篠崎 政久 (SHINOZAKI, Masahisa); 〒2128585 神奈川県川崎市幸区堀川町7番地34 東芝デジタルソリューションズ株式会社内 Kanagawa (JP). 酒主 佳祐 (SAKANUSHI, Keisuke); 〒2128585 神奈川県川

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

(54) 発明の名称: 情報処理装置、情報処理方法、および情報処理プログラム



(57) Abstract: An information processing device (10) is provided with a dividing unit (20A) and an identification unit (20B). The dividing unit (20A) divides a plurality of sentences included in a processing-target sentence collection (40) into a plurality of groups (G) on the basis of a teacher topic list (16B) that indicates a list of preset teacher topic names. The identification unit (20B) identifies a topic name for each of the groups (G) on the basis of the teacher topic list (16B).

崎市幸区堀川町 7 2 番地 3 4 東芝デジタルソリューションズ株式会社内 Kanagawa (JP).

(74) 代理人: 特許業務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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, 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:

— 国際調査報告 (条約第21条(3))

(57) 要約: 情報処理装置 (10) は、分割部 (20A) と、特定部 (20B) と、を備える。分割部 (20A) は、処理対象の文集合 (40) に含まれる複数の文を、予め設定した教師話題名の一覧を示す教師話題リスト (16B) に基づいて、複数のグループ (G) に分割する。特定部 (20B) は、グループ (G) ごとに、話題教師リスト (16B) に基づいて話題名を特定する。

明 細 書

発明の名称：

情報処理装置、情報処理方法、および情報処理プログラム

技術分野

[0001] 本発明の実施形態は、情報処理装置、情報処理方法、および情報処理プログラムに関する。

背景技術

[0002] 近年の情報システムの高度化に伴い、特許文献、新聞記事、ウェブページ、および、書籍といった文書、並びに、映像および音声などのデータを蓄積することが可能になっている。これらのデータの内容を容易に把握可能とする技術が求められている。

[0003] このような技術の1つとして、テキストデータに含まれる全ての単語について、単位時間あたりの発話密度が所定値を超える単語を、該単位時間における話題語として抽出する技術が提案されている。しかしながら、従来技術では、単位時間ごとに、発話密度の高い単語を話題語として抽出していた。このため、従来技術では、実際の話題より細かい単位で話題名が付与されると共に、実際の話題より細かい意味を示す単語が話題名として付与されていた。

先行技術文献

特許文献

[0004] 特許文献1：特開2008-9552号公報

発明の概要

発明が解決しようとする課題

[0005] 本発明は、処理対象の文集合に対して、実際の話題により近い区間ごとに、実際の話題により近い話題名を特定することができる、情報処理装置、情報処理方法、および情報処理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

[0006] 実施形態の情報処理装置は、分割部と、特定部と、を備える。分割部は、処理対象の文集合に含まれる複数の文を、予め設定した教師話題名の一覧を示す教師話題リストに基づいて、複数のグループに分割する。特定部は、グループごとに、教師話題リストに基づいて、話題名を特定する。

図面の簡単な説明

[0007] [図1]図1は、本実施の形態の情報処理システムの構成の一例を示す模式図である。

[図2]図2は、録音時に端末に表示される表示画面の一例を示す模式図である。

[図3]図3は、履歴管理DBのデータ構成の一例を示す模式図である。

[図4]図4は、教師話題リストのデータ構成の一例を示す模式図である。

[図5]図5は、話題ベクトルの算出の一例を示す説明図である。

[図6]図6は、話題ベクトルの算出の一例を示す模式図である。

[図7]図7は、類似度算出の一例を示す模式図である。

[図8]図8は、グループへの分割の一例の説明図である。

[図9A]図9Aは、表示制御部が表示する表示画面の一例を示す模式図である。

[図9B]図9Bは、表示制御部が表示する表示画面の一例を示す模式図である。

[図10]図10は、情報処理装置で実行される情報処理の手順の一例を示すフローチャートである。

[図11]図11は、分割処理の手順の一例を示すフローチャートである。

[図12]図12は、話題名特定処理の手順の一例を示すフローチャートである。

[図13]図13は、本実施の形態の情報処理装置のハードウェア構成例を示す説明図である。

発明を実施するための形態

[0008] 以下に添付図面を参照して、情報処理装置、情報処理方法、および情報処理プログラムの、実施の形態を詳細に説明する。

[0009] 図1は、本実施の形態の情報処理システム1000の構成の一例を示す模式図である。

[0010] 本実施の形態の情報処理システム1000は、情報処理装置10と、端末12と、認識装置14と、記憶装置16と、を備える。情報処理装置10、端末12、認識装置14、および記憶装置16は、ネットワーク18を介して通信可能に接続されている。

[0011] ネットワーク18は、例えば、LAN（ローカルエリアネットワーク）、および、インターネットなどである。ネットワーク18の形態はこれらに限られず、任意のネットワーク形態とすることができる。

[0012] なお、情報処理装置10、端末12、認識装置14、および記憶装置16の少なくとも1つは、無線または有線によりネットワーク18に接続されている。無線により接続する場合、例えば、3G回線、無線LAN、Bluetooth（登録商標）などを用いればよい。

[0013] また、図1には、情報処理装置10、端末12、認識装置14、および記憶装置16を、別体として構成した場合を示した。しかし、情報処理装置10、端末12、認識装置14、および記憶装置16の少なくとも2つ以上を、1つの装置として構成してもよい。

[0014] また、情報処理装置10、端末12、認識装置14、および記憶装置16の各々の機能の少なくとも一部が、他の装置に備えられていてもよい。また、情報処理装置10、端末12、認識装置14、および記憶装置16の各々の機能の一部または全部を、1つの装置内に備えた構成としてもよい。

[0015] まず、端末12について説明する。端末12は、ユーザによって操作される端末装置である。端末12は、例えば、スマートフォン、タブレット、および、PC（パーソナルコンピュータ）などである。

[0016] 端末12は、制御部30と、音声入力部32と、操作パネル34と、通信部36と、を備える。音声入力部32、操作パネル34、および通信部36

は、制御部 30 に接続されている。

[0017] 音声入力部 32 は、ユーザにより発話された音声等を入力する。音声入力部 32 は、例えば、マイクロフォンである。

[0018] 操作パネル 34 は、ユーザの操作に応じた各種の操作入力を受付けるとともに、各種の情報や各種の画像を表示する。

[0019] 本実施の形態では、操作パネル 34 は、操作部 34 A と、表示部 34 B と、の双方を一体的に備えたタッチパネルである場合を説明する。操作部 34 A は、各種操作入力の受付けを行う受付機能を実現する。表示部 34 B は、各種情報の表示を行う表示機能を実現する。なお、操作パネル 34 の構成は、このような構成に限定されない。例えば、操作パネル 34 は、操作部 34 A と、表示部 34 B と、を別体として構成してもよい。

[0020] 通信部 36 は、ネットワーク 18 を介して他の装置と通信する。

[0021] 制御部 30 は、端末 12 を制御する。制御部 30 は、受付部 30 A と、表示制御部 30 B と、を有する。受付部 30 A および表示制御部 30 B の一部または全ては、例えば、CPU (Central Processing Unit) などの処理装置にプログラムを実行させること（すなわちソフトウェア）により実現してもよいし、IC (Integrated Circuit) などのハードウェアにより実現してもよいし、併用して実現してもよい。

[0022] 受付部 30 A は、ユーザによる操作部 34 A の操作指示によって示される信号を受付ける。表示制御部 30 B は、表示部 34 B への表示処理を制御する。例えば、表示制御部 30 B は、表示画面などの各種の画像を表示部 34 B へ表示する制御を行う。

[0023] 例えば、ユーザは、表示部 34 B に表示された表示画面を介して、音声の録音開始や録音終了を指示する。図 2 は、録音時に端末 12 に表示される表示画面 38 の一例を示す模式図である。

[0024] 表示画面 38 は、端末 12 の操作パネル 34 に表示される。表示画面 38 は、音声の認識結果を示す表示領域 38 A と、録音開始ボタン 38 B と、録

音終了ボタン３８Ｃと、再生ボタン３８Ｄと、を含む。

[0025] 例えば、ユーザは、表示画面３８における録音開始ボタン３８Ｂを押下する。すると、受付部３０Ａは、音声入力を開始するように音声入力部３２を制御し、音声入力部３２から音声を受付ける。また、表示制御部３０Ｂは、録音開始ボタン３８Ｂが押下されると、録音開始ボタン３８Ｂに代えて、録音終了ボタン３８Ｃを表示画面３８に表示する。そして、受付部３０Ａは、ユーザによって録音終了ボタン３８Ｃが押下されるまで、音声の受付を継続する。端末１２の通信部３６は、録音開始ボタン３８Ｂが押下されてから録音終了ボタン３８Ｃが押下されるまでの期間に受付けた音声を、１つの音声ファイルとして、認識装置１４へ出力する。表示画面３８の表示領域３８Ａには、例えば、該音声ファイルの認識結果である文が表示される。

[0026] 図１に戻り、説明を続ける。次に、認識装置１４について説明する。認識装置１４は、音声を認識し、認識結果を示すテキストデータを出力する。例えば、認識装置１４は、端末１２の音声入力部３２で録音された音声の音声ファイルを受信する。そして、認識装置１４は、音声ファイルを解析することで、音声ファイルに含まれる音声をテキスト（文）に変換し、テキストデータを得る。

[0027] テキストデータは、複数の文の集合体である。複数の文の集合体を、文集合と称して説明する。言い換えると、文集合は、複数の文を含む。認識装置１４によって得られた１または複数の文は、記憶装置１６に記憶される。

[0028] 次に、記憶装置１６について説明する。記憶装置１６は、各種の情報を記憶する記憶部を備える。記憶装置１６は、ファイルサーバ、クラウドサーバ、と称される場合がある。記憶装置１６の記憶部は、例えば、ＨＤＤ（Hard Disk Drive）、光ディスク、メモリカード、ＲＡＭ（Random Access Memory）などの一般的に利用されているあらゆる記憶媒体により構成することができる。

[0029] 本実施の形態では、記憶装置１６は、履歴管理ＤＢ１６Ａと、教師話題リスト１６Ｂと、を記憶する。

- [0030] 図3は、履歴管理DB16Aのデータ構成の一例を示す模式図である。履歴管理DB16Aは、文集合を管理するためのデータベースである。なお、履歴管理DB16Aのデータ形式は、データベースに限定されない。
- [0031] 例えば、履歴管理DB16Aは、ファイルと、ファイルに含まれる文の文IDと、文と、発話タイミングと、形態素と、を対応づけたものである。なお、履歴管理DB16Aは、少なくとも、後述する情報処理装置10で処理する対象の複数の文を登録したデータベースであればよい。
- [0032] ファイルは、履歴管理DB16Aに登録されている文の認識元のファイルである。すなわち、ファイルは、音声ファイル、文書ファイル、などである。
- [0033] 文IDは、文の識別情報である。発話タイミングは、対応するファイルにおける、文の相対位置を示す情報である。例えば、発話タイミングは、対応する文を示す音声の発話されたタイミングや、対応するファイルにおける位置（例えば、行、列によって示される位置）などを示す。形態素は、対応する文に含まれる形態素である。
- [0034] 例えば、認識装置14は、端末12から受信した音声ファイルと、音声ファイルに含まれる認識した文と、文の発話タイミングと、文に含まれる形態素と、を対応づけて履歴管理DB16Aに記憶する。また、認識装置14は、履歴管理DB16Aに登録されている文の各々に、文IDを付与し、履歴管理DB16Aに登録する。
- [0035] なお、履歴管理DB16Aに文書ファイルを登録する場合、例えば、記憶装置16の制御部が、該文書ファイルを公知の方法により解析し、文ID、発話タイミング、文、および形態素、を対応づけて履歴管理DB16Aに登録すればよい。この場合、発話タイミングには、各文の各々の、文書ファイルにおける位置を用いればよい。
- [0036] なお、形態素の解析方法には、公知の形態素解析を用いればよい。また、形態素の解析および履歴管理DB16Aへの登録は、認識装置14、記憶装置16、および後述する情報処理装置10の何れで行ってもよい。

[0037] 次に、記憶装置 16 に記憶される、教師話題リスト 16 B について説明する。図 4 は、教師話題リスト 16 B のデータ構成の一例を示す模式図である。

[0038] 教師話題リスト 16 B は、予め設定した教師話題名の一覧を示す。教師話題名とは、予め想定および設定した話題名を示す。教師話題名は、スーパーバイズド・トピック・ネームと称してもよい。

[0039] 例えば、教師話題リスト 16 B は、教師話題名と、文・単語と、を対応づけたものである。教師話題名に対応する、文・単語とは、対応する教師話題名の話題に属する文（テキスト）や単語を示す。すなわち、教師話題リスト 16 B における教師話題名の各々は、対応する文・単語によって規定される話題の名称であり、予め設定および想定されたものである。

[0040] なお、本明細書において、文集合とは、複数の文の集合である。すなわち、上述したように、文集合は、複数の文を含む。文は、文法上の言語単位であり、文字で書かれたまとまった一連の単語の集合である。文は、複数の形態素から構成される。単語は、形態素の内の自由形態素である。自由形態素は、他の形態素と連結せずに、単体で単語となれる形態素である。教師話題名および話題名は、単語の 1 種であり、教師話題名や話題名として選択された単語であるといえる。

[0041] 教師話題リスト 16 B は、予め設定される。教師話題リスト 16 B は、ユーザによる操作指示などによって変更可能としてもよい。本実施の形態では、記憶装置 16 は、1 つの教師話題リスト 16 B を予め記憶する場合を、一例として説明する。しかし、記憶装置 16 は、予め定めた分野ごとに、1 または複数の教師話題リスト 16 B を予め記憶してもよい。この場合、記憶装置 16 は、分野の識別情報（分野 ID）と、教師話題リスト 16 B と、を対応づけて記憶すればよい。

[0042] 図 1 に戻り説明を続ける。次に、情報処理装置 10 について説明する。情報処理装置 10 は、処理対象の文集合に対して、分割したグループごとに話題名を特定する（詳細後述）。

- [0043] 図1に示すように、情報処理装置10は、制御部20と、通信部22と、を備える。通信部22は、ネットワーク18を介して、端末12、認識装置14、および記憶装置16と通信する。通信部22は、制御部20に接続されている。
- [0044] 制御部20は、情報処理装置10を制御する。制御部20は、分割部20Aと、特定部20Bと、表示制御部20Cと、受付部20Dと、階層表示制御部20Eと、を備える。分割部20Aは、分類部20Fと、確率算出部20Gと、類似度算出部20Hと、グループ分割部20Iと、を含む。特定部20Bは、判断部20Jと、第1特定部20Kと、第2特定部20Lと、を含む。
- [0045] 分割部20A、特定部20B、表示制御部20C、受付部20D、階層表示制御部20E、分類部20F、確率算出部20G、類似度算出部20Hと、グループ分割部20Iと、判断部20J、第1特定部20Kと、および第2特定部20Lの一部または全ては、例えば、CPUなどの処理装置にプログラムを実行させること（すなわちソフトウェア）により実現してもよいし、ICなどのハードウェアにより実現してもよいし、併用して実現してもよい。
- [0046] 分割部20Aは、処理対象の文集合に含まれる複数の文を、教師話題リスト16Bに基づいて、複数のグループに分割する。本実施の形態では、分割部20Aは、処理対象の文集合に含まれる複数の文を、発話タイミング順に並べる。そして、分割部20Aは、教師話題リスト16Bに基づいて、少なくとも1つの文間に区切り位置を設ける。これによって、分割部20Aは、処理対象の文集合を、該区切り位置で区切られた複数のグループに分割する。
- [0047] 以下、詳細に説明する。まず、分割部20Aは、処理対象の文集合を取得する。処理対象の文集合とは、処理対象の範囲の複数の文を示す。言い換えると、処理対象の文集合とは、話題名の特定対象の、文集合である。処理対象の範囲は、処理対象の文を特定可能な情報によって示されればよい。例え

ば、処理対象の範囲は、発話タイミングの特定の期間や、文書ファイルにおける特定の領域、などを示す情報である。更に詳細には、処理対象の範囲は、特定の年、特定の月、特定の年の特定の月、特定の日、特定の年月日、特定の日における特定の時間帯、などによって表される。

[0048] 分割部20Aは、処理対象の文を履歴管理DB16Aから読取ること、処理対象の文集合を取得する。例えば、分割部20Aは、ある期間内に発話された複数の文を履歴管理DB16Aから読み取ること、処理対象の文集合を取得する。

[0049] なお、処理対象の範囲は、予め設定してもよいし、ユーザによる操作指示によって変更可能としてもよい。また、端末12のユーザが端末12の操作パネル34を操作することで、処理対象の範囲を入力してもよい。この場合、情報処理装置10の分割部20Aは、端末12から、処理対象の範囲を受け付け、該範囲の複数の文を履歴管理DB16Aから読取ること、処理対象の文集合を取得すればよい。

[0050] 次に、分割部20Aは、処理対象の文集合に含まれる複数の文を、複数のグループに分割する。

[0051] 分割部20Aは、処理対象の文集合に含まれる文を、教師話題リスト16Bに基づいて、複数のグループに分割する。教師話題リスト16Bを用いることによって、分割部20Aは、教師話題リスト16Bに登録されている教師話題名に応じた話題の区間ごとに、文集合に含まれる文を分割することができる。

[0052] 本実施の形態では、分割部20Aは、分類部20Fと、確率算出部20Gと、類似度算出部20Hと、グループ分割部20Iと、を含む。

[0053] 分類部20Fは、処理対象の文集合に含まれる単語を、複数の話題カテゴリに分類する。分類部20Fは、処理対象の文集合を構成する複数の文の各々に含まれる全ての単語を抽出し、抽出した複数の単語を、複数の話題カテゴリに分類する。

[0054] 話題カテゴリとは、同じ話題である可能性の高いカテゴリである。分類部

20Fは、公知の方法を用いて、処理対象の文集合に含まれる複数の単語を、複数の話題カテゴリに分類する。例えば、分類部20Fは、LDA (Latent Dirichlet Allocation) を用いて、処理対象の文集合に含まれる複数の単語を、複数の話題カテゴリに分類する。

[0055] LDAは、文集合から指定した個数の話題カテゴリを抽出し、各話題カテゴリにおける、文集合に含まれる単語の各々の属する確率と、文集合に含まれる各文における各話題カテゴリの属する確率と、の各々を導出する手法である（例えば、参考文献1参照）（参考文献1：David M. Blei, Andrew Y. Ng, Michael I. Jordan, Latent Dirichlet Allocation, Journal of machine Learning research, Vol. 3, pp. 993-1022, 2003.）。

[0056] 本実施の形態では、分類部20Fは、話題カテゴリごとに、識別情報（話題カテゴリ1～話題カテゴリM（Mは2以上の整数））を付与し、後の処理に用いるものとして説明する。

[0057] 確率算出部20Gは、処理対象の文集合に含まれる文ごとに、話題ベクトルを算出する。話題ベクトルは、文集合に含まれる文における、分類部20Fで分類された複数の話題カテゴリの各々が属する確率を示す。本実施の形態では、確率算出部20Gは、教師話題リスト16Bに基づいて、複数の話題カテゴリの各々が属する確率を算出する際に、教師話題リスト16Bに示される教師話題名に対応する話題カテゴリが属する確率を、実際の確率より大きくなるように、文ごとに話題ベクトルを算出する。実際の確率とは、話題ベクトルにおける、複数の話題カテゴリの各々の属する確率であって、何等かの要素（ここでは教師話題名）に応じた重み付けなどのなされていない状態の確率である。

[0058] この大きくする割合（すなわち、重み付け値）は、予め設定すればよい。また、この重み付け値は、ユーザによる端末12の操作指示などによって、適宜調整可能としてもよい。この重み付け値を調整することで、教師話題リスト16Bに登録されている教師話題名に応じた話題の区間に沿った、グループへの分割が可能となる。

- [0059] 図5は、話題ベクトルの算出の一例を示す説明図である。図5中の一番上の図は、処理対象の文集合40の一例を示す模式図である。例えば、文集合40が、文1D1～文1Dm（mは2以上の整数）に対応する複数の文を含んでいたと仮定する。
- [0060] この場合、分類部20Fは、例えば、図5中の2番目の図に示すように、文集合40に含まれる複数の単語を、話題カテゴリ1～話題カテゴリMの複数のカテゴリに分類する。単語所属確率43は、処理対象の文集合に含まれる複数の単語の各々が属する確率を示す。
- [0061] そして、確率算出部20Gは、分類部20Fで分類した各話題カテゴリ1～話題カテゴリMの各々について、単語所属確率43を算出する（図5中の2番目の図参照）。単語所属確率43は、処理対象の文集合に含まれる複数の単語の各々が属する確率を示す。
- [0062] 確率算出部20Gは、LDAを用いて、話題カテゴリ1～話題カテゴリMの各々について、単語所属確率43を算出すればよい。
- [0063] 図5中の2番目の図には、一例として、確率算出部20Gが、話題カテゴリ1～話題カテゴリ4の各々について、単語所属確率43₁～単語所属確率43₄を算出した形態を示した。なお、確率算出部20Gは、他の話題カテゴリMの各々についても同様にして、単語所属確率43を算出する。
- [0064] そして、確率算出部20Gは、処理対象の文集合40に含まれる文ごとに、話題ベクトル41を算出する（図5中の3番目の図参照）。
- [0065] 話題ベクトル41は、上述したように、処理対象の文集合に含まれる文における、分類部20Fで分類された複数の話題カテゴリの各々が属する確率を示す。詳細には、図5中の3番目の図に示すように、各文に対応する話題ベクトル41は、分類部20Fによって分類された話題カテゴリの数であるM個の要素からなる。そして、話題ベクトル41に示される、M個の要素の各々の値は、対応する話題カテゴリの属する確率を示す値である。
- [0066] 確率算出部20Gは、LDAを用いて、処理対象の文集合に含まれる文ごとに、話題ベクトル41を算出すればよい。詳細には、確率算出部20Gは

、話題カテゴリの各々の単語所属確率43に基づいて、LDAを用いて、処理対象の文集合に含まれる文ごとに、話題ベクトル41を算出する。

[0067] 図5中の3番目の図には、文ID1～文ID3によって識別される各々の文に対応する、話題ベクトル41（話題ベクトル41a～話題ベクトル41c）を一例として示した。なお、確率算出部20Gは、処理対象の文集合に含まれる他の文の各々についても同様に、話題ベクトル41を算出する。

[0068] ここで、確率算出部20Gは、分類部20Fによって分類された複数の話題カテゴリ1～話題カテゴリMの内、教師話題リスト16Bに登録されている教師話題名に対応する話題カテゴリについては、確率が実際の確率より大きくなるように、話題ベクトル41を算出する。なお、実際の確率には、単語所属確率43から算出した確率を用いればよい。

[0069] 例えば、話題カテゴリ1～話題カテゴリMの内、話題カテゴリ1が、教師話題リスト16Bにおける教師話題名「認定子供園」に対応すると仮定する。また、話題カテゴリ2が、教師話題リスト16Bにおける、教師話題名「子育て支援」に対応すると仮定する。また、話題カテゴリ3が、教師話題リスト16Bにおける、教師話題名「出産」に対応すると仮定する。なお、話題カテゴリ4については、教師話題リスト16Bにおける教師話題の何れにも対応しないと仮定する。

[0070] この場合、確率算出部20Gは、話題ベクトル41（話題ベクトル41a～話題ベクトル41c）の各々における、教師話題名「認定子供園」、「子育て支援」、「出産」に対応する話題カテゴリ1～話題カテゴリ3の各々の属する確率が大きくなるように、これらの確率を補正する（図5中の4番目の図参照）。そして、補正後の話題ベクトル41（話題ベクトル41a'～話題ベクトル41c'）の各々を、各文に対応する話題ベクトル41として算出する。

[0071] なお、話題カテゴリ1～話題カテゴリMが、教師話題リスト16Bにおける教師話題名の何れに対応するかは、公知の方法で導出すればよい。例えば、確率算出部20Gは、複数の話題カテゴリ1～話題カテゴリMの内、教師

話題リスト 16B における文・単語に示される単語の属する確率が閾値以上の話題カテゴリを、該単語に対応する教師話題名に対応すると特定する。一方、確率算出部 20G は、複数の話題カテゴリ 1～M の内、教師話題リスト 16B における文・単語に示される単語の属する確率が閾値未満の話題カテゴリについては、対応する教師話題名が無いものとする。

[0072] なお、確率算出部 20G は、各話題カテゴリ 1～話題カテゴリ M の各々の単語所属確率 43 から、教師話題名に基づいて話題カテゴリの属する確率を補正した話題ベクトル 41 を、直接算出してもよい。

[0073] この場合、確率算出部 20G は、まず、話題カテゴリの各々の単語所属確率 43 を算出した段階で（図 5 中の 2 番目の図参照）、各話題カテゴリの各々に対応する教師話題名を特定する。話題カテゴリの各々に対応する教師話題名の特定は、上述した方法で行えばよい。また、上述したように、対応する教師話題名の無い話題カテゴリも存在する場合がある。

[0074] そして、確率算出部 20G は、話題カテゴリの各々の単語所属確率 43 から、LDA を用いて各文の話題ベクトル 41 を算出するときに、教師話題名の特定された話題カテゴリについては、実際の確率より大きくなるように、属する確率を算出すればよい。

[0075] 例えば、確率算出部 20G は、LDA-Z（参考文献 2 を参照）（参考文献 2 : David Andrzejewski, Xiaojin Zhu, Latent Dirichlet Allocation with topic-in-set knowledge, In Proceedings of the NAACL HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing pp.43-38, 2009.）を用いる。LDA-Z を用いることによって、確率算出部 20G は、話題カテゴリ 1～話題カテゴリ M の各々に対応する教師話題名の特定と、図 5 中の 4 番目の図に示す補正後の話題ベクトル 41（話題ベクトル 41a' ～ 41c'）の算出と、を行うことができる。

[0076] この場合、確率算出部 20G は、例えば、図 5 中の 2 番目の図に示す各話題カテゴリの単語所属確率 43 から、図 5 中の 3 番目の図に示す補正前の話題ベクトル 41 を経由することなく、図 5 中の 4 番目の図に示す教師話題名

に基づいて確率を補正した話題ベクトル41（話題ベクトル41a'～話題ベクトル41c'）を直接算出することができる。なお、情報処理装置10の制御部20は、以降の処理では、教師話題名に基づいて確率を補正した話題ベクトル41を用いる。

[0077] 図1に戻り説明を続ける。類似度算出部20Hは、処理対象の文集合について、該文集合における位置が隣接する文間の各々における、複数の話題カテゴリの各々の属する確率の類似度を算出する。すなわち、類似度算出部20Hは、処理対象の文集合について、位置が隣接する文間の各々における、話題ベクトル41の類似度を算出する。

[0078] 文間とは、文集合における位置が隣接する文と文との間を示す。

[0079] 文集合における位置が隣接する文、とは、処理対象の文集合における相対位置が隣接する事を示す。すなわち、文集合における位置が隣接する、とは、発話タイミングが隣接する事を示す。発話タイミングは、上述したように、文に対応する音声の発話されたタイミングや、対応するファイルにおける記載位置を示す。このため、発話タイミングが音声の発話されたタイミングを示す場合、位置が隣接する文とは、発話されたタイミングが隣接する文である。また、発話タイミングが、対応するファイルにおける記載位置を示す場合、位置が隣接する文とは、ファイルにおける記載位置が書字方向に沿って隣接する文を示す。

[0080] なお、類似度算出部20Hは、文間の各々における、話題ベクトル41の類似度を算出してもよいが、処理対象の文集合に含まれる文を、予め定めた結合数の文の群（文群と称する）ごとにまとめた上で、文群の間の各々における話題ベクトル41の類似度を算出してもよい。

[0081] 文群は、処理対象の文集合に含まれる文を、文の書字方向または文の発話タイミングの昇順を示す方向に沿って、予め定めた結合数ごとにまとめたものである。

[0082] この場合、確率算出部20Gは、処理対象の文集合に含まれる文を、予め定めた結合数の文群ごとにまとめた上で、文群ごとに話題ベクトル41を算

出する。すなわち、確率算出部20Gは、文群に含まれる複数の文の各々に対応する話題ベクトル41を1つにまとめて、1つの文群に対して1つの話題ベクトル41を算出する。

[0083] なお、結合数は、1以上の整数であればよい。結合数は、例えば、予めユーザが設定する。なお、確率算出部20Gは、処理対象の文集合に含まれる文の数に応じて、結合数を算出してもよい。例えば、確率算出部20Gは、1以上で且つ処理対象の文集合に含まれる文の数未満の整数の範囲内で、結合数を算出する。具体的には、例えば、確率算出部20Gは、結合数の最大値を、処理対象の文集合に含まれる文の数の $1/2$ の数未満とする。そして、確率算出部20Gは、例えば、処理対象の文集合に含まれる文の数が多いほど、より大きい数の結合数を算出すればよい。

[0084] そして、確率算出部20Gは、文群に対応する話題ベクトル41を算出する。すなわち、確率算出部20Gは、1つの文群に対して1つの話題ベクトル41を算出する。

[0085] 図6は、結合数“3”であるときの、文群ごとの話題ベクトル41の算出の一例を示す模式図である。

[0086] 図6の左の図は、処理対象の文集合40の一例を示す模式図である。例えば、処理対象の文集合40が、文ID1～文IDmの各々によって識別される文から構成されていると仮定する。なお、処理対象の文集合40に含まれる文は、発話タイミングに沿って昇順に（認定元のファイルが文書ファイルである場合は書字方向に沿って）配列されているものとする。

[0087] この場合、まず、確率算出部20Gは、文集合40を構成する複数の文を、結合数‘3’に応じて、3つの文ごとにまとめることで、文群を特定する。例えば、類似度算出部20Hは、文ID1～文ID3の各々によって識別される3つの文を、1つの文群として扱う。確率算出部20Gは、文ID4～文IDmの各々によって識別される文についても、結合数‘3’に応じて、3つの文ごとにまとめ、文群を特定する。

[0088] そして、確率算出部20Gは、各文群に属する文の各々の話題ベクトル4

1 を 1 つにまとめ、文群に対応する話題ベクトル 4 1 を算出する。

[0089] 詳細には、確率算出部 2 0 G は、文 1 D 1 ~ 文 1 D 3 の 3 つの文からなる文群に属する文の各々の、話題ベクトル 4 1 (話題ベクトル 4 1 a' ~ 話題ベクトル 4 1 c') を特定する。この話題ベクトル 4 1 a' ~ 話題ベクトル 4 1 c' は、図 5 中の 4 番目の図に示す、教師話題名に基づいて確率を補正した話題ベクトル 4 1 (話題ベクトル 4 1 a' ~ 話題ベクトル 4 1 c') である。

[0090] 次に、確率算出部 2 0 G は、文群に属する文の各々に対応する話題ベクトル 4 1 (話題ベクトル 4 1 a' ~ 話題ベクトル 4 1 c') について、複数の話題カテゴリの各々の属する確率を、該話題ベクトルにおける話題カテゴリの属する確率の最大値で除算する。これによって、確率算出部 2 0 G は、文群に属する文の各々に対応する話題ベクトル 4 1 に示される話題カテゴリの各々の属する確率を正規化した、話題ベクトル 4 1 (話題ベクトル 4 1 a' ~ 話題ベクトル 4 1 c') を導出する。

[0091] そして、確率算出部 2 0 G は、文群に属する文の各々の、正規化した話題ベクトル 4 1 (話題ベクトル 4 1 a' ~ 話題ベクトル 4 1 c') について、話題カテゴリごとに、属する確率の合計値を算出する。これによって、確率算出部 2 0 G は、文群に対応する話題ベクトル 4 1 を算出する。図 6 には、文 1 D 1 ~ 文 1 D 3 の文からなる文群に対応する話題ベクトル 4 1 として、話題ベクトル 4 1 A を示した。

[0092] 確率算出部 2 0 G は、他の文群についても同様にして、文群ごとに対応する話題ベクトル 4 1 を算出する。

[0093] 次に、類似度算出部 2 0 H は、処理対象の文集合に含まれる文間の各々における、話題ベクトル 4 1 の類似度を算出する。本実施の形態では、類似度算出部 2 0 H は、処理対象の文集合に含まれる、上記結合数ごとにまとめた文の文群間の各々について、話題ベクトル 4 1 の類似度を算出する。

[0094] 図 7 は、類似度算出の一例を示す模式図である。図 7 中の左の図は、処理対象の文集合 4 0 の一例を示す模式図である。図 7 中の左の図は、図 6 中の

左の図と同様である。

[0095] そして、上記処理によって、確率算出部20Gが、文ID1～文ID3の文の文群に対応する話題ベクトル41として、話題ベクトル41Aを算出したと仮定する。また、確率算出部20Gが、文ID4～文ID6の文の文群に対応する話題ベクトル41として、話題ベクトル41Bを算出したと仮定する。また、確率算出部20Gは、文ID7～文ID9に対応する話題ベクトル41として、話題ベクトル41Cを算出したと仮定する。なお、確率算出部20Gは、他の文群に対しても同様に、対応する話題ベクトル41を算出している。

[0096] この場合、類似度算出部20Hは、隣接する文群の話題ベクトル41の類似度を算出する。具体的には、類似度算出部20Hは、話題ベクトル41Aと話題ベクトル41Bとの類似度を算出する。同様に、類似度算出部20Hは、話題ベクトル41Bと話題ベクトル41Cとの類似度や、他の隣接する文群の話題ベクトル41の類似度を算出する。

[0097] 類似度算出部20Hは、隣接する2つの文群にそれぞれ対応する2つの話題ベクトル41について、同じ話題カテゴリの属する確率をベクトルとして扱ったコサイン距離（コサイン類似度）を、類似度として算出する。

[0098] なお、上述したように、類似度算出部20Hは、処理対象の文集合について、該文集合における位置が隣接する文間の各々における、話題ベクトル41の類似度を算出してもよい。但し、文集合における発話タイミングに沿って配列された文を結合数に応じてまとめた文群の間の各々における、話題ベクトル41の類似度を算出することが好ましい。

[0099] ここで、通常の会話や文書においては、同じ話題が所定期間継続することが多い。このため、文間における話題ベクトル41の類似度は、文群の間における話題ベクトル41の類似度に比べて小さくなることが多い。詳細は後述するが、本実施の形態では、グループ分割部20Iは、類似度に応じて、文集合に含まれる複数の文を、複数のグループに分割する。

[0100] このため、類似度算出部20Hが、結合数に応じてまとめた文群の間にお

ける話題ベクトル41の類似度を算出することで、グループ分割部201は、文集合を、より多くの文を含むグループに分割することができる。また、情報処理装置10における情報速度の高速化および処理負荷の軽減を図ることができる。

[0101] 図1に戻り説明を続ける。次に、グループ分割部201について説明する。グループ分割部201は、処理対象の文集合に含まれる複数の文を、類似度に応じて複数のグループに分割する。

[0102] 詳細には、グループ分割部201は、類似度算出部20Hで算出された複数の類似度の内、低い順に予め定めた数（ $K-1$ 個）の類似度を示す文間を区切り位置として、処理対象の文集合に含まれる複数の文を、複数の（ K 個）のグループに分割する。

[0103] 区切り位置の数（ $K-1$ ）個は、予め定めればよい。例えば、この区切り位置の数（ $K-1$ ）個は、上記結合数と同様に、処理対象の文集合に含まれる文の数に応じて予め定めればよい。また、この区切り位置の数は、ユーザによる端末12の操作指示などに応じて、適宜変更可能としてもよい。

[0104] また、グループ分割部201は、処理対象の文集合に含まれる文の数に応じて、区切り位置の数（ $K-1$ 個）を算出し、グループへの分割に用いてもよい。この場合、グループ分割部201は、上記結合数の算出と同様にして、区切り位置の数を算出すればよい。

[0105] なお、グループ分割部201は、類似度算出部20Hで算出された類似度の内、閾値以下の類似度を示す文間を区切り位置として、文集合に含まれる複数の文を、複数のグループに分割してもよい。この閾値は、予め設定すればよい。また、この閾値は、ユーザによる端末12の操作指示などに応じて、適宜変更可能としてもよい。

[0106] 図8は、グループGへの分割の一例の説明図である。例えば、処理対象の文集合40が、図8に示す文集合40であったと仮定する。なお、図8に示す文集合40は、図6中の左の図および図7中の左の図と同様である。

[0107] そして、グループ分割部201が、類似度に応じて、文ID6と文ID7

の文間、文 I D 9 と文 I D 1 0 との文間、などを区切り位置 5 0（区切り位置 5 0 A、区切り位置 5 0 B）として特定したと仮定する。この場合、グループ分割部 2 0 I は、文集合 4 0 に含まれる複数の文を、区切り位置 5 0 を境界として、複数のグループ G（グループ G 1 ～グループ G n（n は 2 以上の整数））に分割する。

[0108] 上記処理によって、分割部 2 0 A は、処理対象の文集合 4 0 に含まれる複数の文を、複数のグループ G に分割する。

[0109] 図 1 に戻り説明を続ける。次に、特定部 2 0 B について説明する。

[0110] 特定部 2 0 B は、分割部 2 0 A によって分割されたグループ G ごとに、グループ G に含まれる文に含まれる単語および教師話題リスト 1 6 B に基づいて、話題名を特定する。

[0111] 本実施の形態では、特定部 2 0 B は、判断部 2 0 J と、第 1 特定部 2 0 K と、第 2 特定部 2 0 L と、を有する。

[0112] 判断部 2 0 J は、グループ G ごとに、属する確率が最大の話題カテゴリが、教師話題リスト 1 6 B における何れかの教師話題名に対応するか否かを判断する。

[0113] なお、グループ G には、複数の文が含まれる。このため、判断部 2 0 J は、類似度算出部 2 0 H による文群に対応する話題ベクトル 4 1 の算出時と同様にして、グループ G に含まれる複数の文の各々の話題ベクトル 4 1 を、グループ G ごとに 1 つにまとめる。すなわち、判断部 2 0 J は、1 つのグループ G に対して 1 つの話題ベクトル 4 1 が得られるように、グループ G に属する複数の文の各々の話題ベクトル 4 1 を 1 つにまとめる。

[0114] そして、判断部 2 0 J は、各グループ G に対応する、該まとめた 1 つの話題ベクトル 4 1 について、含まれる話題カテゴリの内、属する確率の最大の 1 つの話題カテゴリが、教師話題名に対応するか否かを判断する。

[0115] なお、判断部 2 0 J は、確率算出部 2 0 G と同様にして、該話題カテゴリが教師話題名の何れかに対応するか否かを判断すればよい。

[0116] ここで、確率算出部 2 0 G が、話題カテゴリ 1 ～話題カテゴリ M の各々に

ついて、対応する教師話題名を特定済である場合がある。この場合、判断部 20J は、グループ G における属する確率が最大の話題カテゴリに対応する話題教師名を、確率算出部 20G から取得すればよい。

[0117] 第1特定部 20K は、判断部 20J によって、対応すると判断されたグループ G について、該グループ G に対応すると判断した教師話題名を話題名として特定する。

[0118] 例えば、グループ G1 における、属する確率が最大の話題カテゴリが話題カテゴリ 1 であったと仮定する。そして、判断部 20J によって、該話題カテゴリ 1 に対応する教師話題名が、教師話題リスト 16B における教師話題名「認定子供園」と判断されたと仮定する。この場合、第1特定部 20K は、グループ G1 の話題名を、「認定子供園」と特定する。また、例えば、グループ G2 における、属する確率が最大の話題カテゴリが話題カテゴリ 2 であったと仮定する。そして、判断部 20J によって、該話題カテゴリ 2 に対応する教師話題名が、教師話題リスト 16B における教師話題名「子育て支援」であったと仮定する。この場合、第1特定部 20K は、グループ G2 の話題名を、「子育て支援」と特定する。

[0119] 一方、第2特定部 20L は、判断部 20J によって対応しないと判断されたグループ G について、処理対象の文集合 40 に含まれる単語の内、該グループ G に属する確率が最大の単語を、話題名として特定する。

[0120] 例えば、グループ G3 における、属する確率が最大の話題カテゴリについては、教師話題名に対応しないと判断されたと仮定する。この場合、第2特定部 20L は、処理対象の文集合 40 に含まれる単語の内、グループ G3 に属する確率が最大の単語（例えば、「あいさつ」を、話題名として特定する。

[0121] 表示制御部 20C は、グループ G ごとに、特定された話題名を表示する。本実施の形態では、表示制御部 20C は、グループ G ごとに、特定された話題名を示す表示画面を生成し、表示部へ表示する。この表示画面を表示する表示部は、情報処理装置 10 に設けられた表示部であってもよいし、端末 1

2の表示部34Bであってもよい。本実施の形態では、一例として、表示制御部20Cは、端末12の表示部34Bに、該表示画面を表示するように制御する場合を説明する。

[0122] 図9A～図9Bは、表示制御部20Cが表示する表示画面60の一例を示す模式図である。

[0123] 例えば、表示制御部20Cは、図9Aに示す表示画面60Aを生成し、表示部34Bへ表示する。表示画面60Aは、タイムライン表示領域Tと、話題名表示領域Sと、を含む。タイムライン表示領域Tは、処理対象の文集合40に含まれる文の発話タイミングに応じたタイムラインの表示領域である。話題名表示領域Sは、グループGごとの話題名の表示領域である。

[0124] 図9Aには、グループG1の話題名「認定子供園」、グループG2の話題名「子育て支援」、グループG3の話題名「あいさつ」の各々を、各グループGに含まれる文の発話タイミングに応じたタイムライン表示領域Tに対応する位置に、話題名表示領域Sに表示した例を示した。このように、表示制御部20Cは、グループGごとに、特定された話題名を表示する。

[0125] なお、表示制御部20Cは、グループGごとに、教師話題名と対応する話題名と、教師話題名と対応しない話題名と、を異なる表示形態で表示することが好ましい。教師話題名と対応する話題名とは、例えば、教師話題名に一致する話題名である。また、教師話題名と対応しない話題名とは、例えば、教師話題名とは不一致の話題名である。具体的には、教師話題名と対応する話題名は、第1特定部20Kによって特定された話題名である。また、教師話題名と対応しない話題名は、第2特定部20Lによって特定された話題名である。

[0126] 異なる表示形態とは、少なくとも視覚的に異なる形態であればよく、例えば、色、形状、および大きさなどが異なることを示す。

[0127] 具体的には、グループG1およびグループG2については、教師話題名と対応する話題名であり、グループG3については、教師話題名と対応しない話題名であったと仮定する。この場合、図9Aに示すように、表示制御部2

〇 C は、教師話題名と対応するグループ G 1 およびグループ G 2 の話題名（「認定子供園」、「子育て支援」）と、教師話題名と対応しないグループ G 3 の話題名「あいさつ」と、について、互いに表示領域の位置および色の異なる表示形態で表示する。

[0128] このため、表示された表示画面 6 〇 を視認するユーザは、処理対象の文集合 4 〇 に含まれる話題を、実際の話題により近い区間ごとに、容易に把握することができる。また、各グループ G に対して特定された話題名は、グループ G に含まれる文の各々の単語を単に用いたものではなく、教師話題リスト 1 6 B に基づいて特定された話題名である。このため、表示画面 6 〇 を視認するユーザは、短時間で、直感的に、処理対象の文集合 4 〇 に含まれる話題を把握することができる。

[0129] また、表示制御部 2 〇 C は、教師話題名と対応するグループ G の話題名と、教師話題名と対応しないグループ G の話題名と、を異なる表示形態で表示する。このため、表示画面 6 〇 を視認するユーザは、表示された話題名によって示される話題が、教師話題リスト 1 6 B に登録されている教師話題名によって示される話題であるか、教師話題名によって示される話題からそれた話題であるか、を容易に把握することができる。

[0130] 図 1 に戻り説明を続ける。ここで、操作パネル 3 4 の表示部 3 4 B に表示された表示画面 6 〇 A について、グループ G ごとに表示された話題名の少なくとも 1 つが指定される場合がある（図 9 A 参照）。

[0131] そこで、情報処理装置 1 〇 は、受付部 2 〇 D および階層表示制御部 2 〇 E を備える。受付部 2 〇 D は、表示された話題名の中の 1 つの指定を受付ける。例えば、端末 1 2 の表示部 3 4 B に、表示画面 6 〇 A（図 9 A 参照）が表示されたと仮定する。そして、端末 1 2 のユーザが、操作部 3 4 A を操作することで、表示画面 6 〇 A における 1 つの話題名を操作指示したと仮定する。すると、端末 1 2 の受付部 3 〇 A は、操作指示された話題名を示す情報を受付け、通信部 3 6 を介して情報処理装置 1 〇 へ送信する。情報処理装置 1 〇 の受付部 2 〇 D は、通信部 2 2 を介して該情報を受信することで、指定さ

れた話題名を受付ける。

[0132] なお、情報処理装置 10 に表示部が設けられ、該表示部に表示画面 60A が表示される場合がある。この場合、情報処理装置 10 の受付部 20D は、情報処理装置 10 を操作するユーザの操作指示によって、表示された話題名の内の 1 つが指定された場合、情報処理装置 10 の操作部から該指定を受付ければよい。

[0133] 階層表示制御部 20E は、受付部 20D で受付けた話題名に対応するグループ G に属する文を、処理対象の文集合として処理するように、分割部 20A および特定部 20B を制御する。

[0134] このため、処理対象の文集合に含まれる文の数が、制御部 20 における前回の処理時に比べて少なくなる。そして、分割部 20A および特定部 20B は、指定された話題名に対応するグループ G に属する複数の文を、新たな処理対象の文集合として用いて、上記と同様の処理を行う。

[0135] ここで、処理対象の文集合として、新たな文の集合を用いると、含まれる文の数や文に含まれる単語の種類の比率などが変わることから、分割部 20A および特定部 20B で用いる単語所属確率 43 や話題ベクトル 41 は、前回の処理時とは異なるものとなる。このため、分割部 20A および特定部 20B によって、指定されたグループ G に属する文の文集合について、前回の処理時より細分化したグループ G への分割や話題名の特定がなされることになる。

[0136] そして、表示制御部 20C が、特定部 20B で新たに特定された話題名について、上記と同様の処理を行うことで、表示部 34B には、指定された話題名に対応するグループ G に属する文を、新たな複数のグループ G に分割した（すなわち、段階的に細分化した）話題名が表示される。

[0137] すなわち、処理対象の文集合として、表示された話題名の内の少なくとも 1 つが指定されることで、前回の処理時より段階的に細分化したグループ G への分割および話題名の特定による、階層処理がなされることとなる。

[0138] 図 9B は、図 9A におけるグループ G1 の話題名「認定子供園」の表示領

域が指示され、該話題名「認定子供園」の指定を受けたときに、表示部 34 B に表示される表示画面 60 B の一例を示す模式図である。

[0139] 階層表示制御部 20 E が、受付部 20 D で受付けた話題名に対応するグループ G に属する文を、処理対象の文集合として処理するように、分割部 20 A および特定部 20 B を制御することによって、表示部 34 B には、表示画面 60 B が表示される。表示画面 60 B における話題名表示領域 S に示されるグループ G は、表示画面 60 A (図 9 A) において指定を受付けたグループ G 1 の話題名「認定子供園」に属する文 (例えば、文 1 D 1 ~ 文 1 D 6 によって識別される文) を、グループ G 1 a とグループ G 1 b に分割したものである。そして、表示画面 60 B には、グループ G 1 a に対して特定された話題名「認定子供園」と、グループ G 1 b に対して特定された話題名「認定」と、が表示される。また、教師話題名と対応するか対応しないかにより、互いに異なる表示形態で話題名が表示される。

[0140] 次に、情報処理装置 10 で実行される情報処理の手順の一例を説明する。

[0141] 図 10 は、情報処理装置 10 で実行される情報処理の手順の一例を示す、フローチャートである。

[0142] まず、分類部 20 F が、処理対象の文集合 40 を取得する (ステップ S 100)。例えば、分類部 20 F は、ユーザによる操作指示によって指定された処理対象の範囲の複数の文を、履歴管理 DB 16 A から読取る。これによって、分類部 20 F は、処理対象の文集合 40 を取得する。

[0143] 次に、分割部 20 A が、取得した処理対象の文集合 40 に含まれる複数の文を複数のグループ G に分割する、分割処理を実行する (ステップ S 102)。ステップ S 102 の分割処理の詳細は、後述する。

[0144] 次に、特定部 20 B が、ステップ S 102 で分割されたグループ G ごとに話題名を特定する、話題名特定処理を実行する (ステップ S 104)。ステップ S 104 の話題名特定処理の詳細は、後述する。

[0145] 次に、表示制御部 20 C が、ステップ S 102 で分割されたグループ G ごとに、ステップ S 104 で特定された話題名を表示するための表示画面 60

を生成する（ステップS 1 0 8）（図9参照）。そして、表示制御部2 0 Cは、該表示画面6 0を表示するように表示制御する（ステップS 1 0 8）。例えば、表示制御部2 0 Cは、ステップS 1 0 6で生成した表示画面6 0を、端末1 2の表示部3 4 Bに表示するように、表示制御する。

[0146] 次に、受付部2 0 Dが、ステップS 1 0 8で表示した表示画面6 0における、話題名の1つの指定を受付けたか否かを判断する（ステップS 1 1 0）。ステップS 1 1 0で否定判断すると（ステップS 1 1 0：N o）、本ルーチンを終了する。

[0147] 一方、ステップS 1 1 0で肯定判断すると（ステップS 1 1 0：Y e s）、ステップS 1 1 2へ進む。ステップS 1 1 2では、階層表示制御部2 0 Eが、受付けた話題名に対応するグループGに属する文の群を特定する（ステップS 1 1 2）。

[0148] そして、階層表示制御部2 0 Eは、特定した文の群を、新たな処理対象の文集合4 0として、分割部2 0 Aへ出力する（ステップS 1 1 4）。これにより、分割部2 0 Aは、処理対象の文集合4 0を新たに取得する（ステップS 1 1 6）。そして、上記ステップS 1 0 2へ戻る。

[0149] 次に、分割部2 0 Aによる、ステップS 1 0 2の分割処理の手順の一例を説明する。図1 1は、ステップS 1 0 2（図1 0参照）の分割処理の手順の一例を示す、フローチャートである。

[0150] まず、分類部2 0 Fが、記憶装置1 6から教師話題リスト1 6 Bを読取る（ステップS 2 0 0）。記憶装置1 6に複数の教師話題リスト1 6 Bが記憶されている場合には、分類部2 0 Fは、これらの複数の教師話題リスト1 6 Bを記憶装置1 6から読取ってもよい。また、分類部2 0 Fは、処理対象の文集合4 0を解析し、処理に用いる教師話題リスト1 6 Bを特定した上で、特定した教師話題リスト1 6 Bを読取ってもよい。

[0151] 次に、分類部2 0 Fは、処理対象の文集合4 0を構成する複数の文の各々に含まれる全ての単語を抽出する（ステップS 2 0 2）。そして、分類部2 0 Fは、ステップS 2 0 2で抽出した単語を、複数の話題カテゴリに分類す

る（ステップS204）。

[0152] 次に、確率算出部20Gが、処理対象の文集合40について、含まれる文ごとに、ステップS206の処理を実行する。ステップS206では、確率算出部20Gは、文に対応する話題ベクトル41を算出する（ステップS206）。このとき、確率算出部20Gは、上述したように、ステップS200で読取った教師話題リスト16Bに登録されている教師話題名に基づいて確率を補正した、話題ベクトル41を算出する。ステップS206の処理によって、1つの文に対して1つの話題ベクトル41が算出される。

[0153] 次に、確率算出部20Gは、処理対象の文集合に含まれる文を、予め定めた結合数の文の群（文群）ごとにまとめた上で、文群ごとに話題ベクトル41を算出する（ステップS208）。上述したように、確率算出部20Gは、文群に含まれる文の各々に対応する話題ベクトル41について、正規化や話題カテゴリごとの確率の加算などの処理を行うことで、文群に対応する話題ベクトル41を算出する（ステップS208）。ステップS208の処理によって、1つの文群に対して、1つの話題ベクトル41が算出される。

[0154] 次に、類似度算出部20Hは、処理対象の文集合40に含まれる文間の各々における、話題ベクトル41の類似度を算出する。上述したように、本実施の形態では、類似度算出部20Hは、処理対象の文集合40に含まれる、隣接する文群間の各々について、話題ベクトル41の類似度を算出する（ステップS210）。

[0155] 次に、グループ分割部20Iが、処理対象の文集合40に含まれる複数の文を、ステップS210で算出された類似度に応じて、複数のグループGに分割する（ステップS212）。そして、本ルーチンを終了する。

[0156] 次に、特定部20Bが実行する話題名特定処理（ステップS104、図10参照）の手順の一例を説明する。図12は、話題名特定処理の手順の一例を示す、フローチャートである。

[0157] 特定部20Bは、ステップS102（図10および図11）の分割処理によって分割されたグループGごとに、ステップS214～ステップS218

の処理を繰返し実行する。

- [0158] まず、判断部20Jが、該グループGに属する確率が最大の話題カテゴリを特定する（ステップS214）。そして、判断部20Jは、ステップS214で特定した話題カテゴリが、ステップS200（図11参照）で読取った教師話題リスト16Bに登録されている教師話題名に対応するか否かを判断する（ステップS216）。
- [0159] 対応すると判断した場合（ステップS216：Yes）、ステップS218へ進む。ステップS218では、第1特定部20Kが、ステップS216で判断した教師話題名を、処理対象の該グループGの話題名として特定する（ステップS218）。
- [0160] 一方、対応しないと判断した場合（ステップS216：No）、ステップS220へ進む。ステップS220では、第2特定部20Lが、処理対象の文集合40に含まれる単語の内、該グループGに属する確率が最大の単語を、該グループGの話題名として特定する（ステップS220）。
- [0161] そして、特定部20Bは、ステップS102（図10および図11）の分割処理によって分割されたグループGの全てについて、ステップS214～ステップS218の処理を実行すると、ステップS222へ進む。
- [0162] ステップS222では、特定部20Bは、グループGごとにステップS214～ステップS220の処理によって特定された、話題名と、話題名に対応するグループGに属する複数の文と、各文に対応する発話タイミングと、を表示制御部20Cへ出力する（ステップS222）。そして、本ルーチンを終了する。なお、これらの情報を受付けた表示制御部20Cは、受付けたこれらの情報を用いて、表示画面60を生成する（図10のステップS106参照）。
- [0163] 以上説明したように、本実施の形態の情報処理装置10は、分割部20Aと、特定部20Bと、を備える。分割部20Aは、処理対象の文集合40に含まれる複数の文を、予め設定した教師話題名の一覧を示す教師話題リスト16Bに基づいて、複数のグループGに分割する。特定部20Bは、グルー

プGごとに、話題教師リスト16Bに基づいて話題名を特定する。

[0164] 従って、本実施の形態の情報処理装置10は、処理対象の文集合40に対して、実際の話題により近い区間ごとに、実際の話題により近い話題名を特定することができる。

[0165] 次に、本実施の形態の情報処理装置10のハードウェア構成を説明する。

図13は、本実施の形態の情報処理装置10のハードウェア構成例を示す説明図である。

[0166] 本実施の形態の情報処理装置10は、CPU71などの制御装置と、ROM (Read Only Memory) 72やRAM (Random Access Memory) 73などの記憶装置と、ネットワークに接続して通信を行う通信I/F74と、各部を接続するバス75を備えている。

[0167] 本実施の形態の情報処理装置10で実行されるプログラムは、ROM72等に予め組み込まれて提供される。

[0168] 本実施の形態の情報処理装置10で実行されるプログラムは、インストール可能な形式又は実行可能な形式のファイルでCD-ROM (Compact Disk Read Only Memory)、フレキシブルディスク (FD)、CD-R (Compact Disk Recordable)、DVD (Digital Versatile Disk) 等のコンピュータで読み取り可能な記録媒体に記録してコンピュータプログラムプロダクトとして提供されるように構成してもよい。

[0169] さらに、本実施の形態の情報処理装置10で実行されるプログラムを、インターネット等のネットワークに接続されたコンピュータ上に格納し、ネットワーク経由でダウンロードさせることにより提供するように構成してもよい。また、本実施の形態にかかる情報処理装置10で実行されるプログラムをインターネット等のネットワーク経由で提供または配布するように構成してもよい。

[0170] 本実施の形態の情報処理装置10で実行されるプログラムは、コンピュータを上述した情報処理装置10の各部として機能させうる。このコンピュー

タは、CPU 71 がコンピュータ読取可能な記憶媒体からプログラムを主記憶装置上に読み出して実行することができる。

[0171] 上記には、本発明の実施の形態を説明したが、上記実施の形態は、例として提示したものであり、発明の範囲を限定することは意図していない。上記新規な実施の形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。上記実施の形態やその変形は、発明の範囲や要旨に含まれるとともに、請求の範囲に記載された発明とその均等の範囲に含まれる。

請求の範囲

- [請求項1] 処理対象の文集合に含まれる複数の文を、予め設定した教師話題名の一覧を示す教師話題リストに基づいて複数のグループに分割する分割部と、
- 前記グループごとに、前記教師話題リストに基づいて話題名を特定する特定部と、
- を備える情報処理装置。
- [請求項2] 前記分割部は、
- 前記文集合に含まれる単語を複数の話題カテゴリに分類する分類部と、
- 前記教師話題リストに基づいて、前記文集合に含まれる文ごとに、複数の前記話題カテゴリの各々が属する確率を算出する際に、前記教師話題リストに示される前記教師話題名に対応する前記話題カテゴリが属する確率を、実際の確率より大きくなるように算出する確率算出部と、
- 前記文集合における位置が隣接する文間の各々における、複数の前記話題カテゴリの各々が属する確率の、類似度を算出する類似度算出部と、
- 前記文集合に含まれる複数の文を、前記類似度に応じて複数の前記グループに分割するグループ分割部と、
- を備える、請求項1に記載の情報処理装置。
- [請求項3] 前記確率算出部は、
- 前記教師話題リストに基づいて、前記文集合に含まれる文ごとに、複数の前記話題カテゴリの各々が属する確率を示す話題ベクトルを算出し、
- 前記類似度算出部は、前記文間の各々における、前記話題ベクトルの前記類似度を算出する、
- 請求項2に記載の情報処理装置。

- [請求項4] 前記確率算出部は、
複数の前記話題カテゴリの各々における、前記文集合に含まれる複数の単語の各々が属する確率を示す単語所属確率と、前記教師話題リストと、に基づいて、前記文集合に含まれる文ごとに前記話題ベクトルを算出する、
請求項3に記載の情報処理装置。
- [請求項5] 前記類似度算出部は、
処理対象の前記文集合に含まれる文を、予め定めた結合数の文群ごとにまとめ、隣接する前記文群の間の各々における前記話題ベクトルの前記類似度を算出する、
請求項3に記載の情報処理装置。
- [請求項6] 前記グループ分割部は、
前記類似度の内、低い順に予め定めた数の前記類似度を示す前記文間を区切り位置として、前記文集合に含まれる複数の文を、複数の前記グループに分割する、
請求項2に記載の情報処理装置。
- [請求項7] 前記グループ分割部は、
前記類似度の内、閾値以下の前記類似度を示す前記文間を区切り位置として、前記文集合に含まれる複数の文を、複数の前記グループに分割する、
請求項2に記載の情報処理装置。
- [請求項8] 前記特定部は、
前記グループごとに、属する確率が最大の前記話題カテゴリが、前記教師話題リストにおける何れかの前記教師話題名に対応するか否かを判断する判断部と、
対応すると判断された前記グループについて、該グループに対応すると判断された前記教師話題名を前記話題名として特定する第1特定部と、

対応しないと判断された前記グループについて、前記文集合に含まれる単語の内、該グループに属する確率が最大の単語を前記話題名として特定する第2特定部と、

を有する、請求項2に記載の情報処理装置。

[請求項9] 前記グループごとに、特定された前記話題名を表示する表示制御部を更に備える、

請求項1に記載の情報処理装置。

[請求項10] 前記表示制御部は、

前記グループごとに、前記教師話題名と対応する前記話題名と、前記教師話題名と対応しない前記話題名と、を異なる表示形態で表示する、

請求項9に記載の情報処理装置。

[請求項11] 表示された前記話題名の内の1つの指定を受付ける受付部と、

受付けた前記話題名に対応する前記グループに属する文を、処理対象の前記文集合として処理するように、前記分割部および前記特定部を制御する階層表示制御部と、

を更に備える、請求項1に記載の情報処理装置。

[請求項12] 処理対象の文集合に含まれる複数の文を、予め設定した教師話題名の一覧を示す教師話題リストに基づいて複数のグループに分割し、

前記グループごとに、前記教師話題リストに基づいて話題名を特定する、

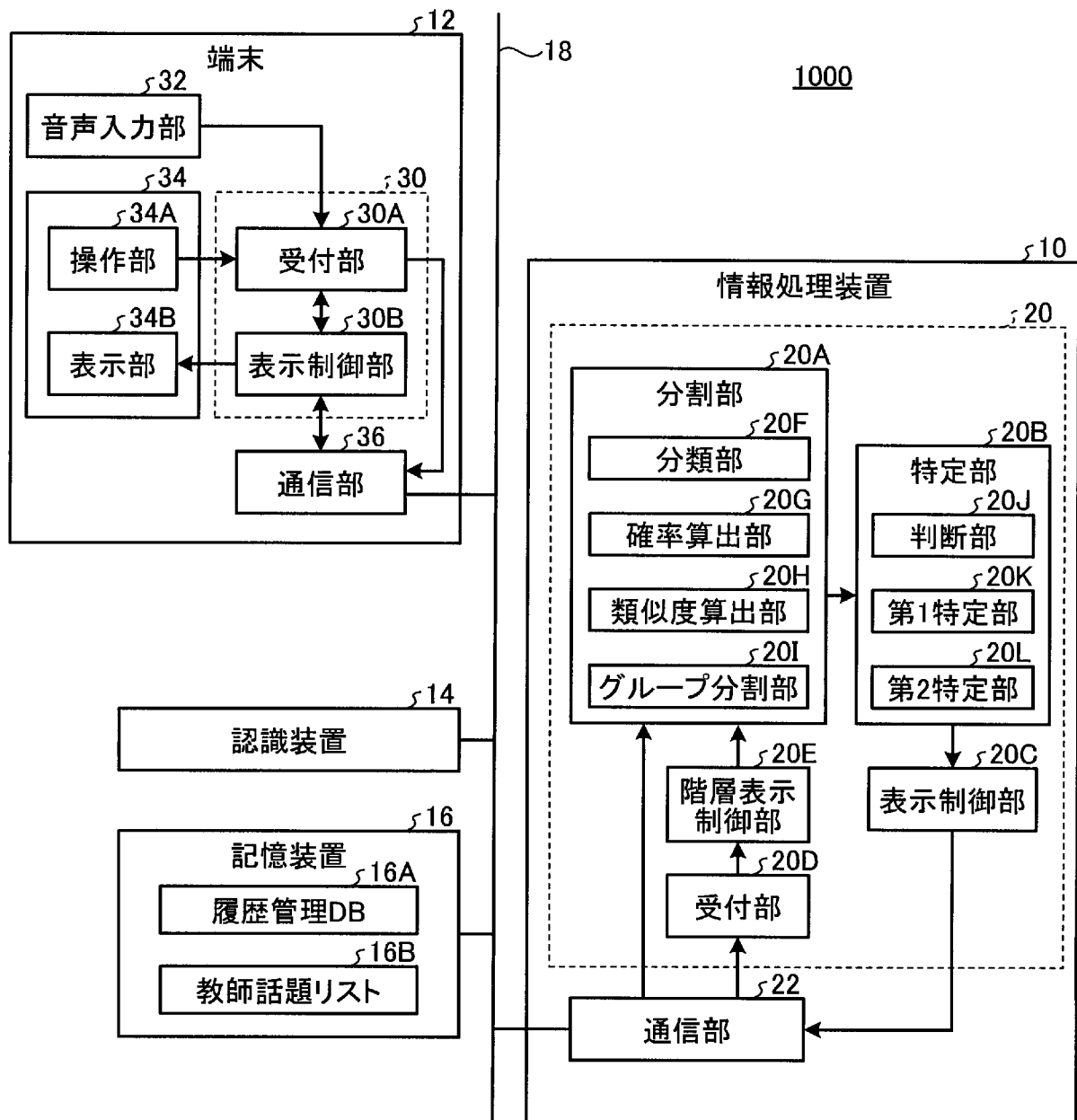
情報処理方法。

[請求項13] 処理対象の文集合に含まれる複数の文を、予め設定した教師話題名の一覧を示す教師話題リストに基づいて複数のグループに分割するステップと、

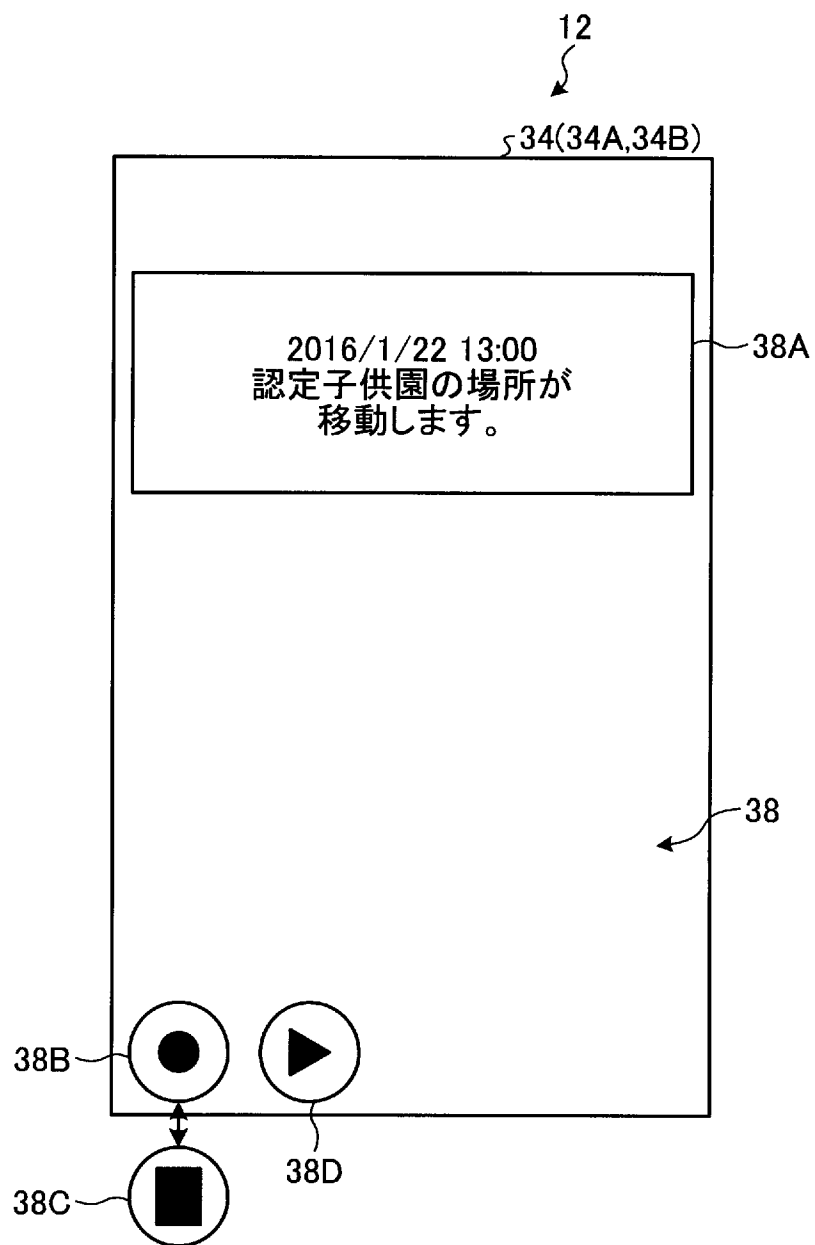
前記グループごとに、前記教師話題リストに基づいて話題名を特定するステップと、

をコンピュータに実行させるための情報処理プログラム。

[図1]



[図2]



[図3]

16A

履歴管理DB

文ID	発話タイミング	ファイル	文	形態素
1	2016/1/22/13:00	AAAA.WAV	認定子供園の場所が移動します。	/認定子供園/の/場所/が/移動/します
2	2016/1/22/13:02		認定子供園について話します。	/認定子供園/に/について/話す/ます
3	2016/1/22/13:05		認定子供園の手続きを透明化していきます。	/認定子供園/の/手続き/を/透明化/して/いき/ま/し/よう
4	2016/1/22/13:06		認定については、どうなっていますか。	/認定/に/について/は/どうなる/います/か
5	2016/1/22/13:07		認定以外の子供園はどうでしょう。	/認定以外/の/子供園/は/どう/でしょう/か
6	2016/1/22/13:08		認定子供園は、手続きが難しいです。	/認定子供園/は/手続き/が/難しい/です
7	2016/1/22/13:10		次に、子育て支援についての質問です。	/子育て支援/に/について/の/質問/です
8	2016/1/22/13:12		子育てを社会全体で応援しましょう。	/子育て/を/社会全体/で/応援/し/ま/し/よう
9	2016/1/22/13:17		子育てにかかる負担軽減のための施策についてです。	/子育て/に/かかる/負担軽減/の/ため/の/施策/に/ついて/です
10	2016/1/22/13:20		ここで、私の自己紹介をさせていただきます。	/それでは/私/の/挨拶/から/始め/させて/いただき/ます
.				
.				
.				

[図4]

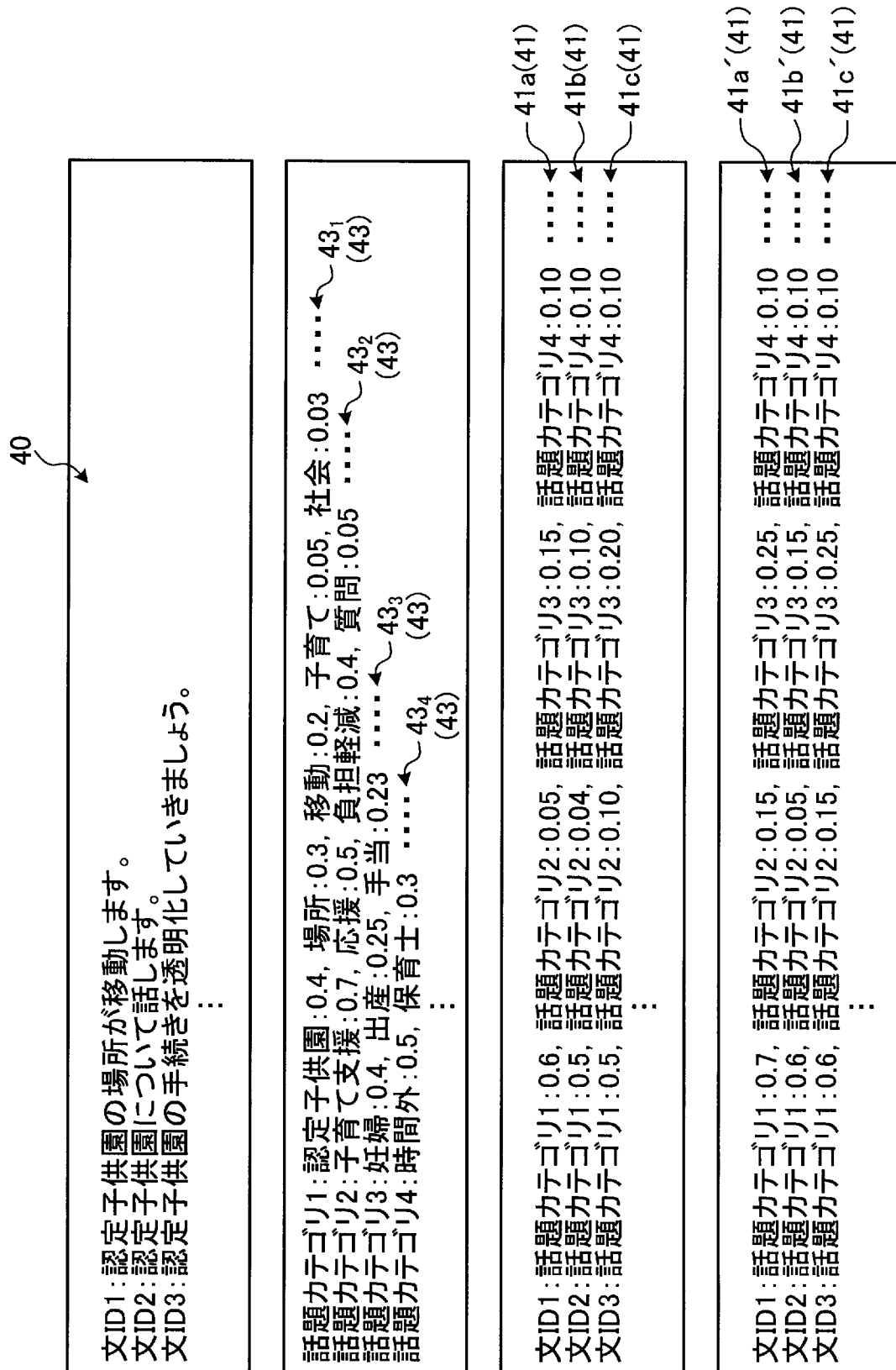
16B



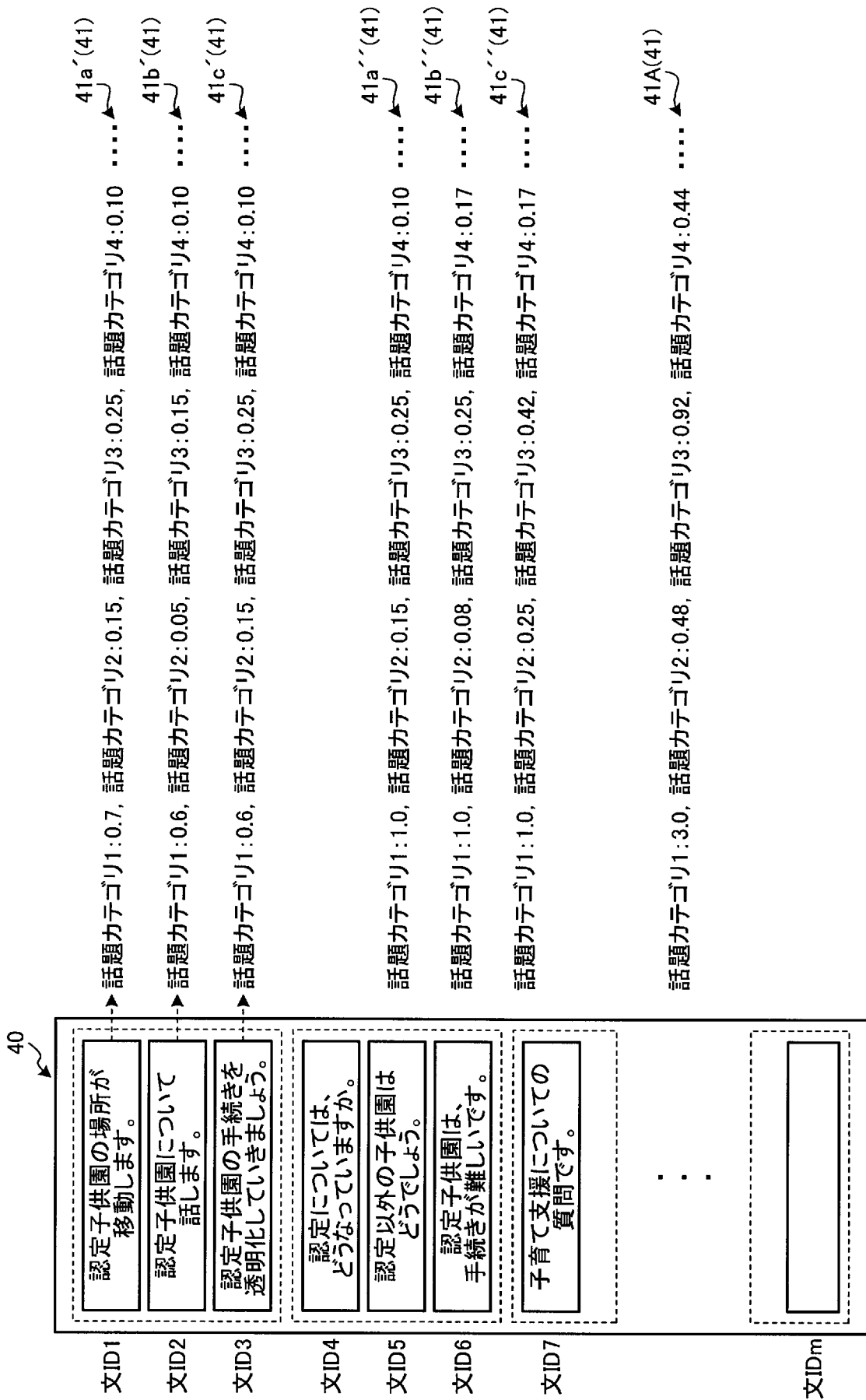
教師話題リスト

教師話題名	文・単語
認定子供園	認定子供園, 手続き, ...
子育て支援	子育て支援について, 子育て, 施策, ...
出産	妊婦, 出産, 出産時, 分娩, ...
	・
	・
	・

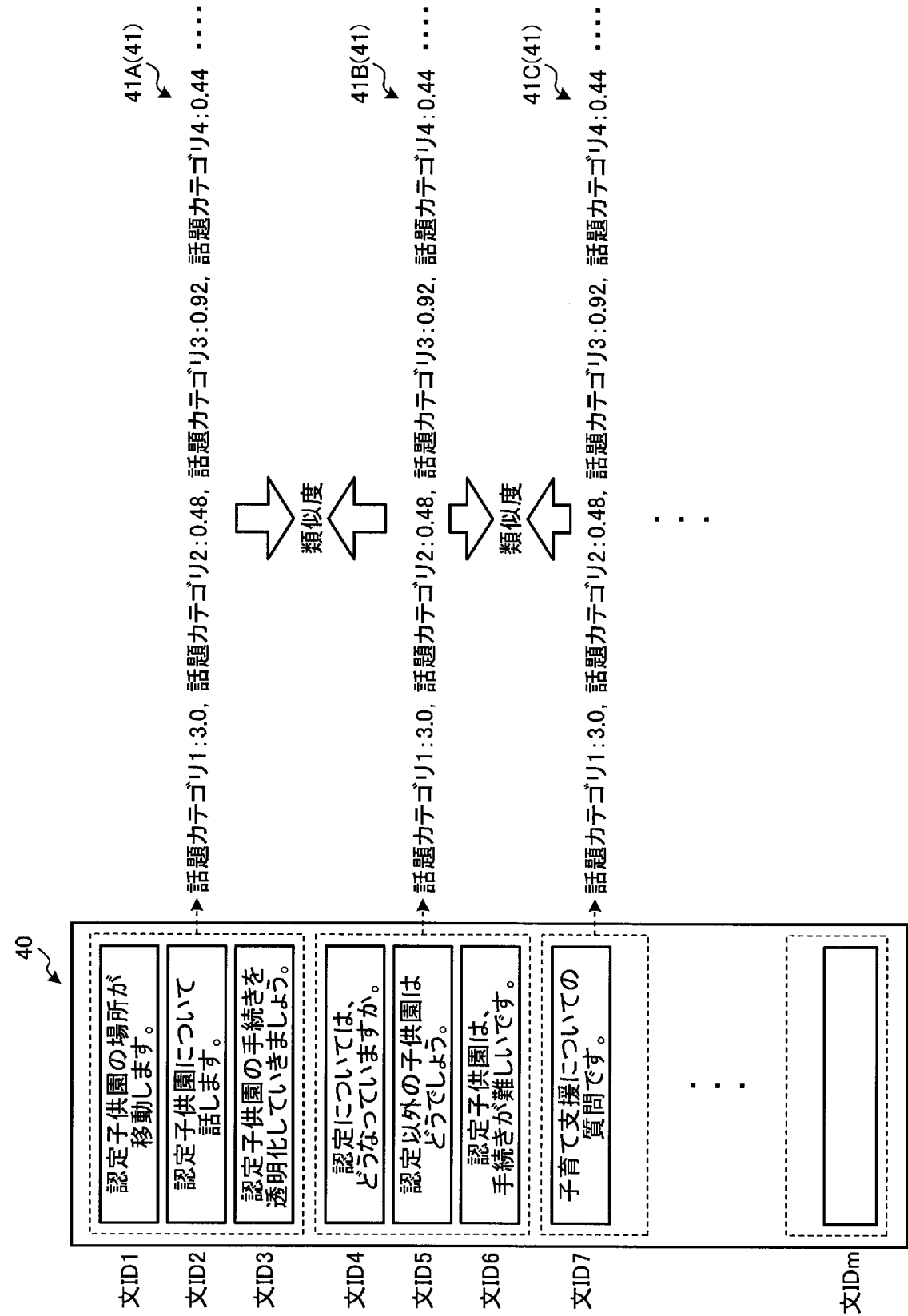
[図5]



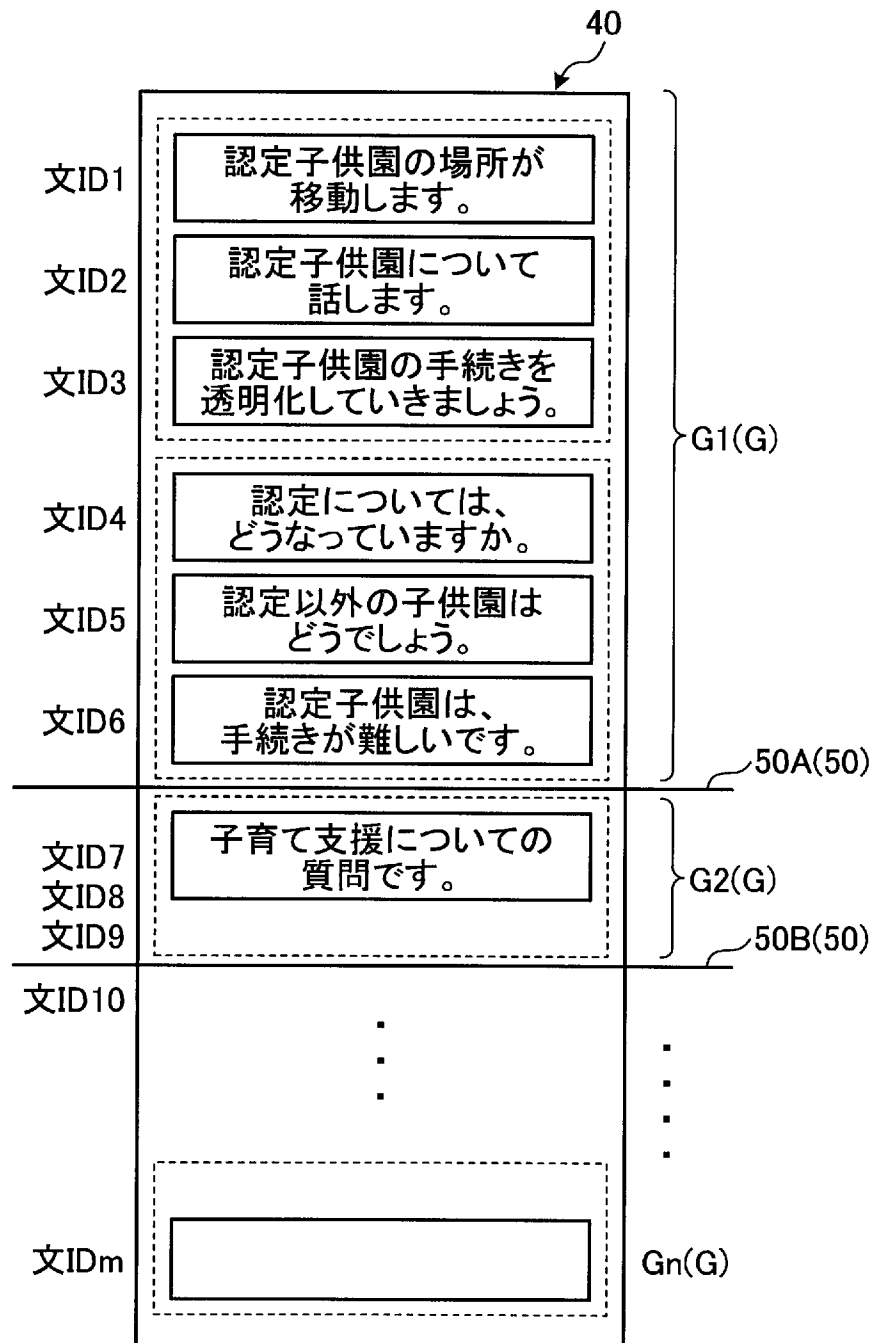
[図6]



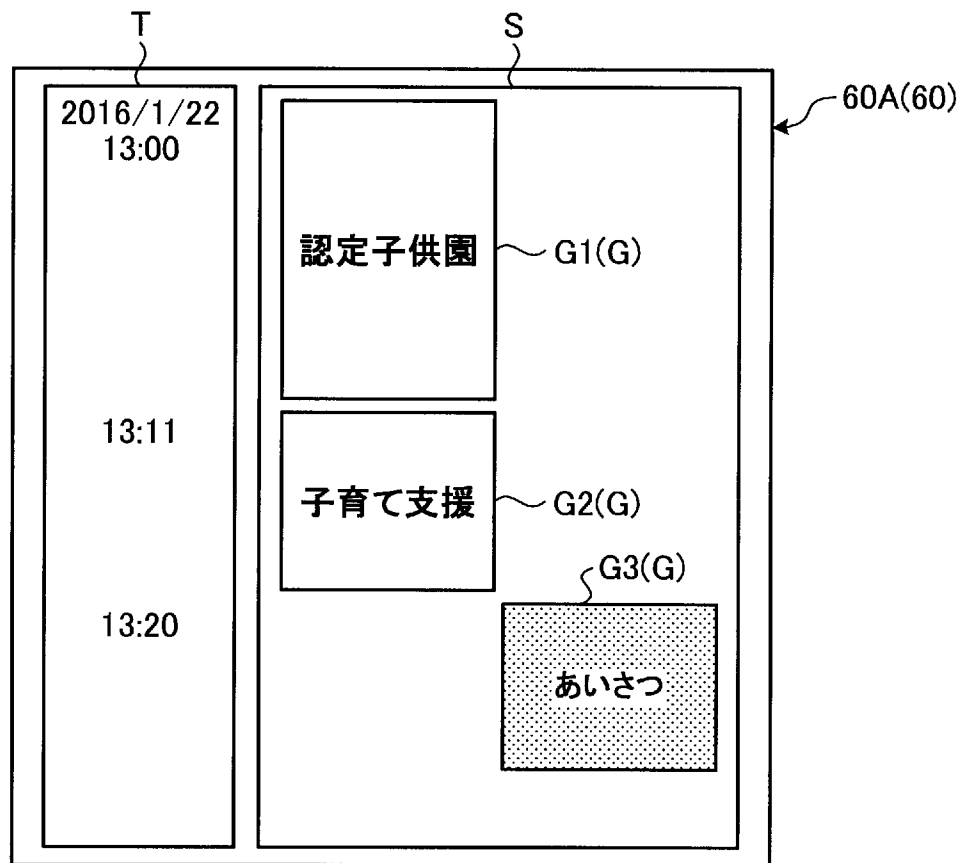
[図7]



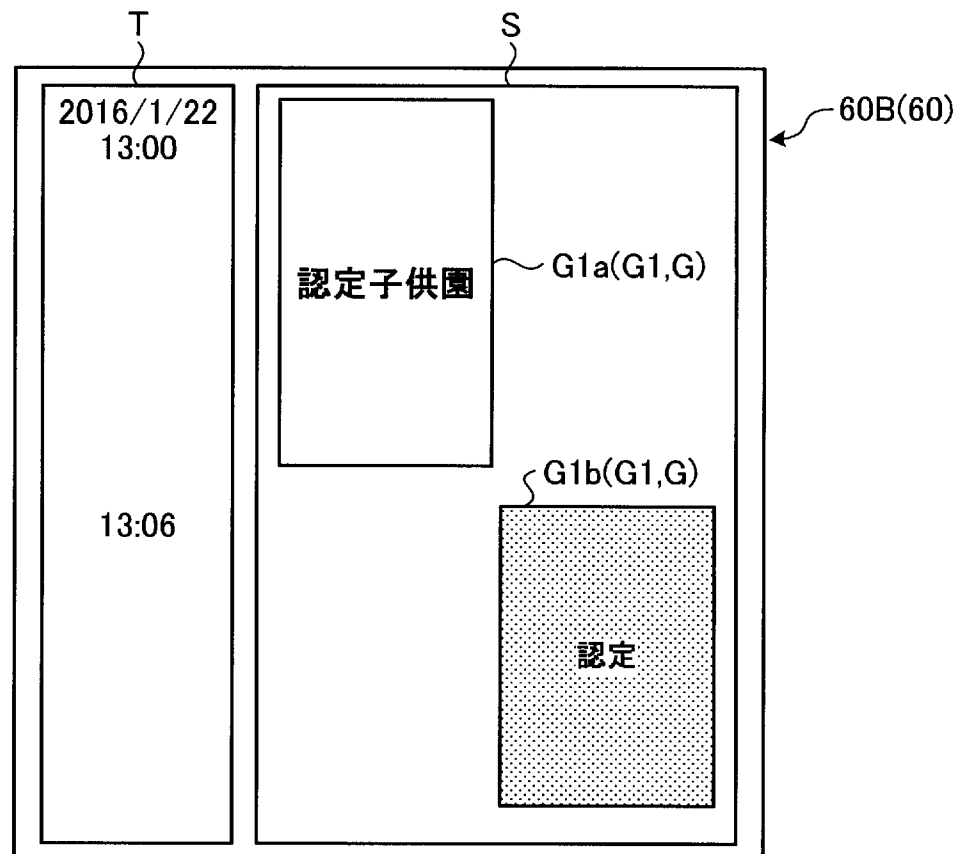
[図8]



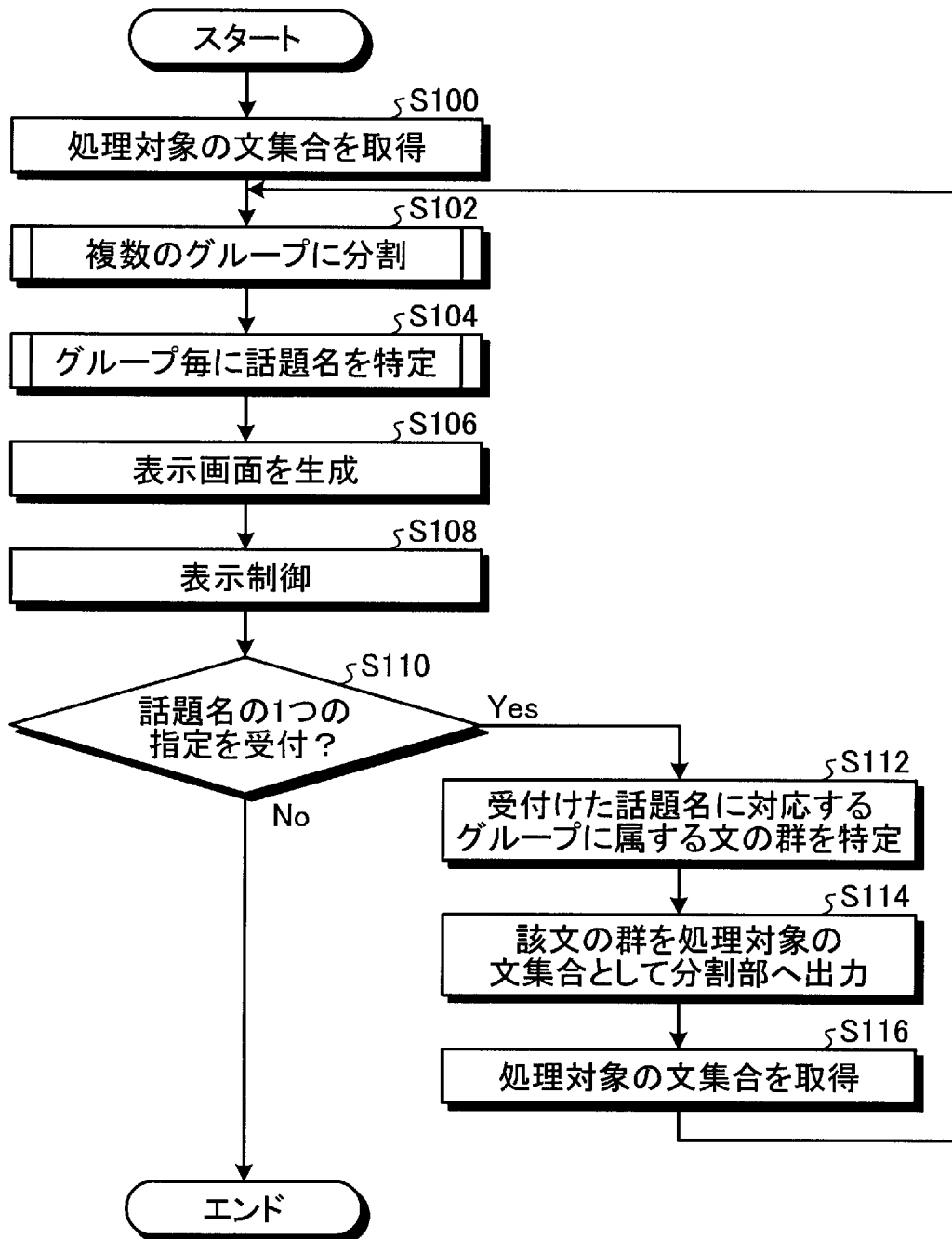
[図9A]



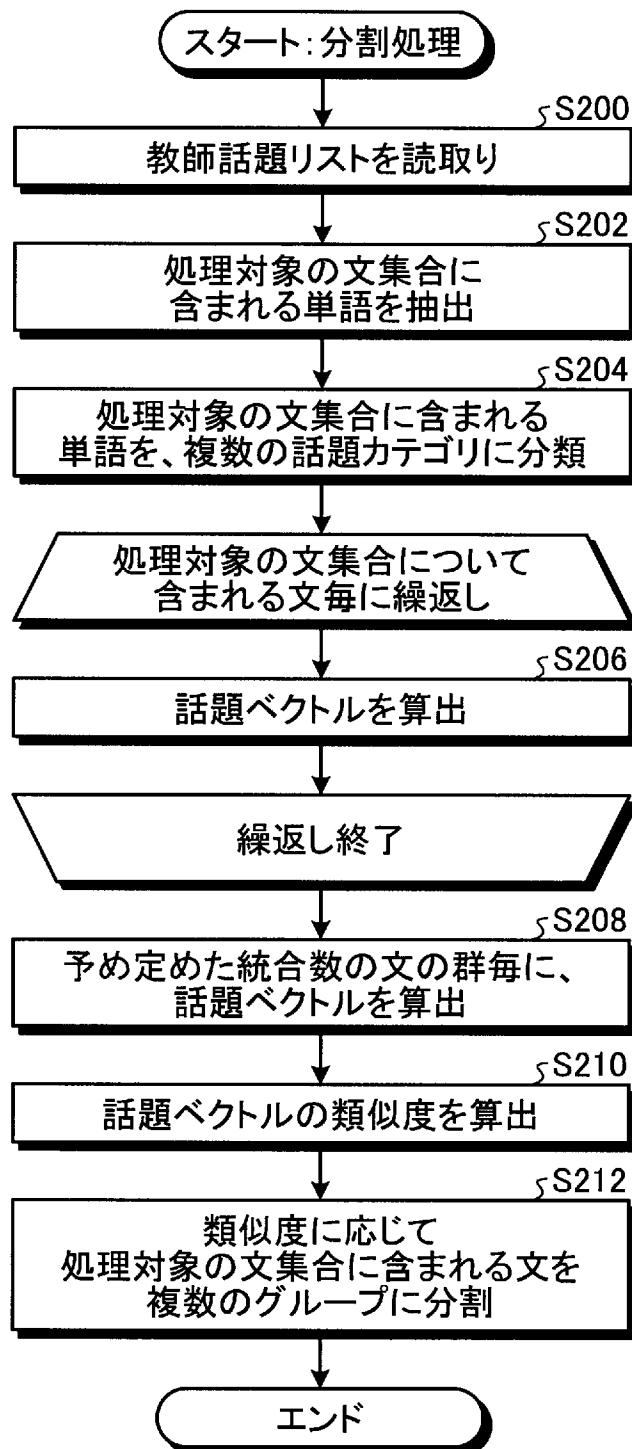
[図9B]



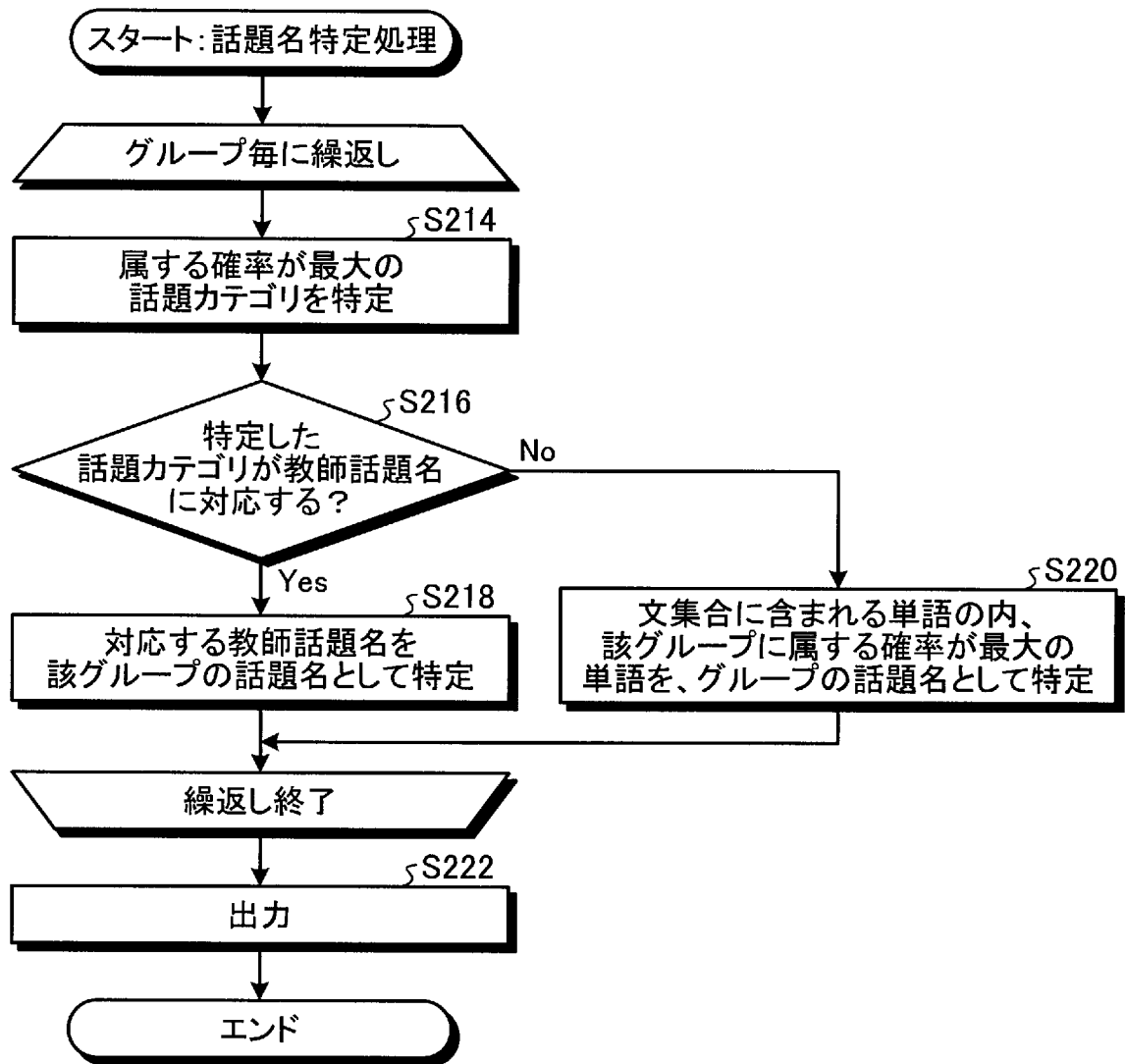
[図10]



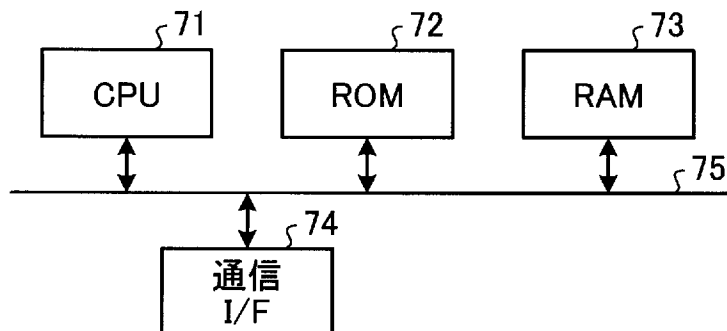
[図11]



[図12]



[図13]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/034583

A. CLASSIFICATION OF SUBJECT MATTER

G06F17/30(2006.01) i, G06F17/27(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, G06F17/27

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	1996-2017
Kokai Jitsuyo Shinan Koho	1971-2017	Toroku Jitsuyo Shinan Koho	1994-2017

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 Y A	JP 2005-196572 A (Japan Science and Technology Agency), 21 July 2005 (21.07.2005), paragraphs [0056] to [0077]; fig. 1 to 5 (Family: none)	1, 9-10, 12-13 11 2-8
Y	JP 2016-103156 A (NTT Comware Corp.), 02 June 2016 (02.06.2016), paragraphs [0049] to [0056]; fig. 6 (Family: none)	11
A	JP 2013-134752 A (Nippon Telegraph and Telephone Corp.), 08 July 2013 (08.07.2013), entire text; all drawings (Family: none)	1-13

☒ 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
25 October 2017 (25.10.17)

Date of mailing of the international search report
07 November 2017 (07.11.17)

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/JP2017/034583

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2016/027364 A1 (Hitachi, Ltd.), 25 February 2016 (25.02.2016), entire text; all drawings (Family: none)	1-13
A	JP 2009-265909 A (Nippon Hoso Kyokai), 12 November 2009 (12.11.2009), entire text; all drawings (Family: none)	1-13
A	JP 2006-350656 A (Nippon Telegraph and Telephone Corp.), 28 December 2006 (28.12.2006), entire text; all drawings (Family: none)	1-13
A	JP 2013-120547 A (Nomura Research Institute, Ltd.), 17 June 2013 (17.06.2013), entire text; all drawings (Family: none)	1-13
A	US 2015/0286627 A1 (ADOBE SYSTEMS INC.), 08 October 2015 (08.10.2015), entire text; all drawings (Family: none)	1-13
A	Katsuhiko ISHIGURO et al., "A survey of the topic models for statistical pattern recognition researchers", IEICE Technical Report, 14 December 2015 (14.12.2015), vol.115, no.388, pages 13 to 24	1-13

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, G06F17/27(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, G06F17/27

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	1971-2017年
日本国実用新案登録公報	1996-2017年
日本国登録実用新案公報	1994-2017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	JP 2005-196572 A（独立行政法人科学技術振興機構）	1, 9-10, 12-13
Y	2005.07.21, 段落[0056]-[0077], [図1]-[図5]	11
A	（ファミリーなし）	2-8
Y	JP 2016-103156 A（エヌ・ティ・ティ・コムウェア株式会社）	11
	2016.06.02, 段落[0049]-[0056], [図6]	
	（ファミリーなし）	

☒ 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

25.10.2017

国際調査報告の発送日

07.11.2017

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）

齊藤 貴孝

5M

4774

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 2013-134752 A (日本電信電話株式会社) 2013. 07. 08, 全文, 全図 (ファミリーなし)	1-13
A	WO 2016/027364 A1 (株式会社日立製作所) 2016. 02. 25, 全文, 全図 (ファミリーなし)	1-13
A	JP 2009-265909 A (日本放送協会) 2009. 11. 12, 全文, 全図 (ファミリーなし)	1-13
A	JP 2006-350656 A (日本電信電話株式会社) 2006. 12. 28, 全文, 全図 (ファミリーなし)	1-13
A	JP 2013-120547 A (株式会社野村総合研究所) 2013. 06. 17, 全文, 全図 (ファミリーなし)	1-13
A	US 2015/0286627 A1 (ADOBE SYSTEMS INCORPORATED) 2015. 10. 08, 全文, 全図 (ファミリーなし)	1-13
A	石黒 勝彦、外 1 名, パターン認識研究者のためのトピックモデリング概観, 電子情報通信学会技術研究報告, 2015. 12. 14, 第 1 1 5 巻, 第 3 8 8 号, p. 1 3 - 2 4	1-13