

(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9月7日(07.09.2018)



(10) 国際公開番号

WO 2018/159010 A1

(51) 国際特許分類:  
G06F 17/30 (2006.01) G06F 21/56 (2013.01)

(21) 国際出願番号: PCT/JP2017/037407

(22) 国際出願日: 2017年10月16日(16.10.2017)

(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 2017-040543 2017年3月3日(03.03.2017) JP

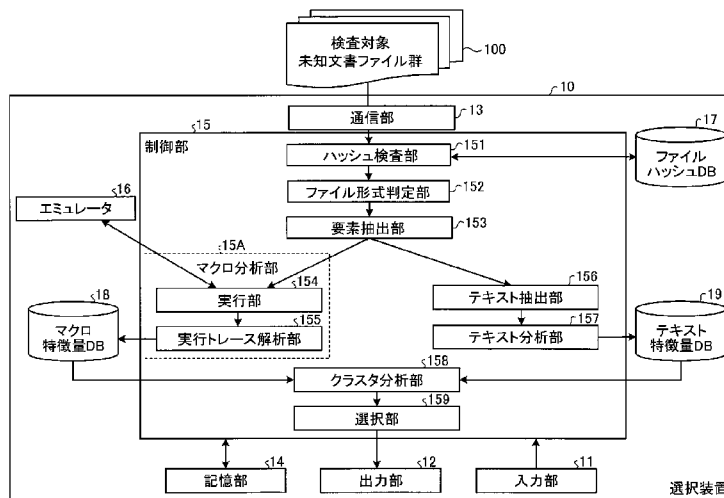
(71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).

(72) 発明者: 碓井 利宣(USUI, Toshinori); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP). 岩村 誠(IWAMURA, Makoto); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP). 矢田 健(YADA, Takeshi); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (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).

(54) Title: SELECTION DEVICE, SELECTION METHOD, AND SELECTION PROGRAM

(54) 発明の名称: 選択装置、選択方法及び選択プログラム



- 10 Selection device
- 11 Input unit
- 12 Output unit
- 13 Communication unit
- 14 Storage unit
- 15 Control unit
- 15A Macro analysis unit
- 16 Emulator
- 17 File hash database
- 18 Macro feature quantity database
- 19 Text feature quantity database
- 100 To-be-verified unknown document file group
- 151 Hash verification unit
- 152 File format determination unit
- 153 Element extraction unit
- 154 Execution unit
- 155 Execution trace analysis unit
- 156 Text extraction unit
- 157 Text analysis unit
- 158 Cluster analysis unit
- 159 Selection unit

(57) Abstract: A selection device (10) is provided with: a macro analysis unit (15A) which acquires a macro feature quantity from a macro within a macro-containing document file; a text analysis unit (157) which acquires a text feature quantity from text within a document file; a cluster analysis unit (158) which executes clustering using the macro feature quantity and the text feature quantity; and a selection unit (159) which, on the basis of a clustering result, selects a document file to be analyzed. The selection device (10) is capable of efficiently and highly accurately selecting a macro-containing document file to be analyzed.

(57) 要約: 選択装置 (10) は、マクロ付き文書ファイル内のマクロからマクロの特徴量を取得するマクロ分析部 (15A) と、文書ファイル内のテキストからテキストの特徴量を取得するテキスト分析部 (157) と、マクロの特徴量とテキストの特徴量とを用いてクラスタリングを実施するクラスタ分析部 (158) と、クラスタ分析結果に基づいて、解析対象の文書ファイルを選択する選択部 (159) と、を有し、解析対象となるマクロ付き文書ファイルを効率的かつ高精度に選択できる。

(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, 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))

## 明 細 書

発明の名称： 選択装置、選択方法及び選択プログラム

### 技術分野

[0001] 本発明は、選択装置、選択方法及び選択プログラムに関する。

### 背景技術

[0002] Microsoft Office（登録商標）向け文書において、マクロを用いた悪性文書ファイル（以下、悪性マクロ付き文書ファイルという。）による脅威が顕在化している。悪性文書ファイルは、内容を閲覧するためにこのファイルをビューアアプリケーション（以下、ビューアという。）で開くと、攻撃が開始され、マルウェアの感染に至るように構成された文書ファイルである。悪性マクロ付き文書ファイルは、このファイルをビューアで開き、マクロが実行されることにより、攻撃が開始される文書ファイルである。

[0003] このマクロを利用できる文書ファイルとしては、Microsoft Office文書があり、ファイル形式の一例としては、DOC、XLS、PPTといったOLE（Object Linking and Embedding）形式の文書ファイルや、DOCM、XLSM、PPTMといった、マクロ有効化されたOffice Open XML（以下、OOXMLという。）形式の文書ファイルがある。

[0004] これらは、Visual Basic for Applications（以下、VBAという。）というプログラミング言語を用いて記述されたマクロを内包できる。VBAを用いたマクロでは、Windowsの提供するAPI（Application Programming Interface）や、ActiveXオブジェクトなどを利用できるため、攻撃者が実現しようとする動作の大半をマクロによって実現できてしまう。そのため、悪性マクロが実行された際にユーザに及ぼす影響度は大きい。

[0005] また、エクスプロイトコードを用いて脆弱性を攻略する悪性文書ファイルでは、攻撃の実現に、悪用可能な脆弱性の発見を必要とし、さらに、攻撃対象の環境がこの脆弱性を有する必要があるという制約が存在する。その一方で、悪性マクロを用いた悪性文書ファイルでは、このような脆弱性の攻略が

必要ないという点で、攻撃の実現が比較的容易である。以上のことから、悪性マクロ付き文書ファイルへの対策の重要性は高い。

[0006] このような悪性マクロ付き文書ファイルからユーザを保護するために、ビューア側で防御機構を提供することが一般的となっている。例えば、前述のOLE形式の文書ファイルに対する最も一般的なビューアであるMicrosoft Officeは、保護ビューと呼ばれるサンドボックスを具備している。保護ビューは、編集機能が無効にした読み取り専用モードであり、保護ビューでは、ユーザが許可しない限り、マクロを実行させることがない。このため、ビューアで悪性マクロ付き文書ファイルを開いた後、直ちにマルウェアに感染してしまう、といった状況を防止できる。このような防御機構の影響もあり、2000年台初頭以降、悪性マクロ付き文書ファイルは、概ね観測されない状態が続いていた。

[0007] ところが、近年では、悪性マクロ付き文書ファイルが再び出現している。これらの悪性マクロ付き文書ファイルは、ユーザに対して保護ビューを無効にするように巧妙に誘導する、いわゆるソーシャルエンジニアリングのためのテキストを文書中に備える。悪性マクロ付き文書ファイルは、これによって、ユーザに自ら保護ビューを解除させ、マクロを実行可能にさせるという手法により、保護ビューの影響を受けることなく攻撃を実現する。保護ビューのような防御機構は、ユーザによって無効にされてしまえば、効果を発揮することができない。このため、このようなソーシャルエンジニアリングの攻撃手法によって、防御機構が十分に機能を果たせなくなってしまう。

[0008] このような悪性マクロ付き文書ファイルは、キャンペーンで多用される。キャンペーンとは、短期間に多数の標的に対して攻撃を実施する一連の攻撃である。悪性なファイルを用いたキャンペーンでは、単一の組織内の多数の標的に対してファイルを送信されることが多い。そのため、組織の入口対策においては、大量の悪性マクロ付き文書ファイルに対して効率的に悪性判定する手法が希求される。

[0009] そこで、悪性マクロ付き文書ファイルによる脅威を検出し、対策をとって

いくための一つの有効な技術として、動的解析に基づく検出技術がある。動的解析は、検査対象の文書ファイルをビューアで開いてマクロを実行し、これが悪質な振る舞いを示すか否かを監視することによって、悪性文書ファイルを検出する手法である。

[0010] マクロへの対策のための動的解析として、従来、以下に挙げる手法が提案されている。例えば、プロセス及びファイルを監視しながら、文書ファイルをビューアアプリケーションで開いてマクロの悪質な動作を実行させることによって、動的に検知する手法が提案されている（例えば、非特許文献1参照）。近年のマクロの多くは、マルウェアのダウンロードやドロップ、実行をおもな悪質な挙動としているため、ファイルの作成やプロセスの生成を監視することによって、その悪質な振る舞いの特徴を捉えることができる。したがって、この非特許文献1記載の手法によれば、悪性マクロ付き文書ファイルを検出できると考えられる。

[0011] また、例えば、文書ファイル内で閉じた動作をするマクロのみの実行を許可し、文書ファイル外の環境への干渉を禁止するという動的な手法が提案されている（例えば、非特許文献2参照）。近年のマクロの持つ悪性動作の多くは、文書ファイル外のシステムの機能を必要とする。このため、この非特許文献2記載の手法によれば、悪性マクロを検出できると考えられる。

[0012] そして、悪性マクロ付き文書ファイルに対する対策のための他の有効な技術として、静的解析に基づく検出技術がある。例えば、文書ファイル内やマクロ内で用いられている単語の出現頻度に基づいて静的に悪性文書を検出する手法が提案されている（例えば、非特許文献3参照）。悪性マクロ付き文書ファイルの用いる単語には傾向があるため、この非特許文献3記載の手法によれば、一定の悪性マクロ、例えば、難読化されていない悪性マクロを、比較的高速で検出できると考えられる。

## 先行技術文献

### 非特許文献

[0013] 非特許文献1：K. Tanuja, CH. Praneeth, Dr D. Haritha, “Dissect

ing Malicious Word, PDF Documents”, International Journal of Science Engineering and Advance Technology, Vol. 3, No. 3, pp. 61–68

非特許文献2: J. Dechaux, E. Filiol, “Proactive defense against malicious documents: formalization, implementation and case studies”, Journal of Computer Virology and Hacking Techniques, Volume 12, Issue 3, pp. 191–202

非特許文献3: Jing-Yao Lin, Hsing-Kuo Pao, “Multi-View Malicious Document Detection”, In Proceedings of the Conference on Technologies and Applications of Artificial Intelligence (TAAI), 2013, pp. 170–175

## 発明の概要

### 発明が解決しようとする課題

[0014] しかしながら、サンドボックスや非特許文献1, 2に記載の手法に代表される動的解析手法では、検査対象の文書ファイルを仮想環境上でマクロを実行させ、悪質な挙動を検出するために、中期的観測（たとえば、一般的なサンドボックスでは5分程度）を実施する必要がある、ファイルの検査に一定の時間を要するという共通した低速性の問題があった。このため、動的解析手法では、キャンペーンと呼ばれる攻撃のように、単一組織に多数のファイルが送信される場合には、適用が困難であった。

[0015] また、非特許文献3に記載の手法では、静的手法のため、難読化された悪性マクロの特徴を捉えられず、文書に出現する単語の特徴のみでは、攻撃者の回避が可能となってしまう、検出精度が低下し得るという問題があった。

[0016] 本発明は、上記に鑑みてなされたものであって、解析対象となるマクロ付き文書ファイルを効率的かつ高精度に選択できる選択装置、選択方法及び選択プログラ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

[0017] 上述した課題を解決し、目的を達成するために、本発明に係る選択装置は

、マクロ付き文書ファイル内のマクロからマクロの特徴量を取得するマクロ分析部と、文書ファイル内のテキストからテキストの特徴量を取得するテキスト分析部と、マクロの特徴量とテキストの特徴量とを用いてクラスタリングを実施するクラスタ分析部と、クラスタリングの結果に基づいて、解析対象の文書ファイルを選択する選択部と、を有することを特徴とする。

## 発明の効果

[0018] 本発明によれば、解析対象となるマクロ文書を効率的かつ高精度に選択できる。

## 図面の簡単な説明

[0019] [図1]図1は、実施の形態に係る選択装置の構成の一例を説明する図である。

[図2]図2は、図1に示すファイルハッシュデータベース(DB)が記憶するデータのデータ形式の一例を示す図である。

[図3]図3は、図1に示すマクロ特徴量DBが記憶するデータのデータ形式の一例を示す図である。

[図4]図4は、図1に示すテキスト特徴量DBが記憶するデータのデータ形式の一例を示す図である。

[図5]図5は、図1に示すエミュレータの構成の一例を説明する図である。

[図6]図6は、図1に示す選択装置の選択処理の処理手順を示すフローチャートである。

[図7]図7は、図6に示すテキスト分析処理の処理手順を示すフローチャートである。

[図8]図8は、図6に示すマクロ分析処理の処理手順を示すフローチャートである。

[図9]図9は、図8に示すエミュレータ実行処理の処理手順を示すフローチャートである。

[図10]図10は、図6に示すクラスタ分析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、図6に示すクラスタ分析処理の他の処理手順を示すフロー

チャートである。

[図12]図12は、図6に示す選択処理の処理手順を示すフローチャートである。

[図13]図13は、プログラムが実行されることにより、選択装置が実現されるコンピュータの一例を示す図である。

### 発明を実施するための形態

[0020] 以下、図面を参照して、本発明の一実施形態を詳細に説明する。なお、この実施形態により本発明が限定されるものではない。また、図面の記載において、同一部分には同一の符号を付して示している。

#### [0021] [実施の形態]

実施の形態に係る選択装置について、選択装置の概略構成、及び、選択装置における処理の流れ及び具体例を説明する。

[0022] 本実施の形態に係る選択装置は、高速な特徴抽出に基づくクラスタリングを実施して類似検体を同一クラスタに分類し、解析対象をクラスタの代表点のみに絞り込むことで、解析のプロセス全体を効率化する。これは、キャンペーンで用いられる多数の悪性マクロ付き文書ファイルは、互いに高い類似性を持つ傾向があるためであり、類似検体をまとめあげることで、類似検体を繰り返し解析することを防ぐ意味を持つ。また、本実施の形態に係る選択装置は、クラスタリングのための特徴抽出には、マクロのエミュレーション実行により、難読化に妨害されない悪性マクロの挙動の特徴を捉えるとともに、文書のテキスト分析により、ソーシャルエンジニアリングのためのテキストの特徴を捉え、攻撃に必須となる両方の特徴を用いた高精度なマルチモーダルクラスタリングを実現する。

[0023] ここで、クラスタリングとは、分類対象の集合を、内的結合（internal cohesion）と外的分離（external isolation）が達成されるような部分集合に分割することである（詳細は、大橋 靖雄，“分類手法概論”，計測と制御，Vol. 24，No. 11，pp. 999—1006，1985参照）。また、マルチモーダルクラスタリングとは、複数の異なる観点からの特徴量に基づいてクラ

スタリングを実現することである。

[0024] [選択装置の構成]

そこで、図1を参照して、実施の形態に係る選択装置の構成について説明する。図1は、実施の形態に係る選択装置の構成の一例を説明する図である。図1に示すように、選択装置10は、入力部11、出力部12、通信部13、記憶部14、制御部15、エミュレータ16、ファイルハッシュデータベース(DB)17、マクロ特徴量DB18、テキスト特徴量DB19を有する。

[0025] 入力部11は、選択装置10の操作者からの各種操作を受け付ける入力インタフェースである。例えば、入力部11は、タッチパネル、音声入力デバイス、キーボードやマウス等の入力デバイスによって構成される。出力部12は、例えば、液晶ディスプレイなどの表示装置、プリンタ等の印刷装置、情報通信装置等によって実現され、選択処理に関する情報を出力する。

[0026] 通信部13は、ネットワーク等を介して接続された他の装置との間で、各種情報を送受信する通信インタフェースである。通信部13は、NIC (Network Interface Card) 等で実現され、LAN (Local Area Network) やインターネットなどの電気通信回線を介した他の装置と制御部15（後述）との間の通信を行う。例えば、通信部13は、検査対象となる未知文書ファイル群（検査対象未知文書ファイル群100）を、制御部15に入力する。また、通信部13は、選択装置10による選択結果情報を、他の装置、或いは、該選択装置10を有する本体装置に送信する。

[0027] 記憶部14は、HDD (Hard Disk Drive)、SSD (Solid State Drive)、光ディスク等の記憶装置である。なお、記憶部14は、RAM (Random Access Memory)、フラッシュメモリ、NVRAM (Non Volatile Static Random Access Memory) 等のデータを書き換え可能な半導体メモリであってもよい。記憶部14は、選択装置10で実行されるOS (Operating System) や各種プログラムを記憶する。さらに、記憶部14は、プログラムの実行で用いられる各種情報を記憶する。

[0028] 制御部15は、選択装置10全体を制御する。制御部15は、例えば、CPU

(Central Processing Unit)、MPU (Micro Processing Unit) 等の電子回路や、ASIC (Application Specific Integrated Circuit)、FPGA (Field Programmable Gate Array) 等の集積回路である。また、制御部 15 は、各種の処理手順を規定したプログラムや制御データを格納するための内部メモリを有し、内部メモリを用いて各処理を実行する。また、制御部 15 は、各種のプログラムが動作することにより各種の処理部として機能する。制御部 15 は、ハッシュ検査部 151、ファイル形式判定部 152、要素抽出部 153、マクロ分析部 15A、テキスト抽出部 156、テキスト分析部 157、クラスタ分析部 158 及び選択部 159 を有する。

[0029] ハッシュ検査部 151 は、入力されるすべてのファイルのハッシュ値を計算する。ハッシュ検査部 151 は、計算の結果、同一のハッシュ値を持つファイルが存在する場合には、重複であるとして 1 つのファイルを残して分析対象から取り除く。ハッシュ検査部 151 は、計算されたハッシュ値を、ファイルハッシュDB 17 (後述) に記憶させる。同一のハッシュ値を持つファイルが存在するか否かの確認は、このファイルハッシュDB 17 の検索により実現する。ハッシュ検査部 151 は、ハッシュ値の計算手法として、暗号学的ハッシュ関数を用いる。この暗号学的ハッシュ関数は、強衝突耐性を保持しているものが望ましい。ハッシュ検査部 151 は、暗号学的ハッシュ関数として、例えば、SHA-2 を用いてもよく、SHA-3 を用いてもよい。

[0030] ファイル形式判定部 152 は、入力されるすべてのファイルに対し、いずれのファイル形式によって構成されているファイルであるかを判定する。ファイル形式判定部 152 は、入力されたファイルについて、このファイルのヘッダやファイル構造の特徴に基づいて、ファイルの形式を判別する。例えば、ファイルの形式として、マクロを内包可能な Microsoft Office 文書である、DOC、XLS、PPT、DOCM、XLSM、PPTM などが想定される。もちろん、マクロを持ち得ないファイルタイプであれば、本選択装置の対象から除外する。

[0031] 要素抽出部 153 は、文書ファイルの構文解析により、文書ファイル中からマクロ及び画像を抽出する。要素抽出部 153 は、抽出したマクロを実行

部 1 5 4（後述）に入力する。また、要素抽出部 1 5 3 は、抽出した画像を、文書ファイルとともにテキスト抽出部 1 5 6（後述）に入力する。要素抽出部 1 5 3 は、マクロの抽出を、文書ファイルのファイルフォーマットの定義に従った構文解析によって実現する。そして、要素抽出部 1 5 3 は、画像の抽出を、前述した構文解析によって実施してもよく、また、ファイルカービングによって実現してもよい。なお、ファイルカービングとは、ファイル中に埋め込まれた他のファイルを、埋め込まれたファイルのマジックナンバーやファイルヘッダ、ファイルフォーマットによって抽出する処理である。

[0032] マクロ分析部 1 5 A は、マクロ付き文書ファイル内のマクロからマクロの特徴量を取得する。マクロ分析部 1 5 A は、実行部 1 5 4 及び実行トレース解析部 1 5 5 を有する。実行部 1 5 4 は、エミュレーション実行の準備をするとともに、エミュレータとのインタフェースとなる。実行部 1 5 4 は、マクロを入力として受け取り、エミュレータ 1 6 にエミュレーション実行させることによって、実行トレースを取得して出力することを目的とする。なお、実行トレースとは、マクロの実行により得られた、マクロのプロファイル情報である。

[0033] 実行部 1 5 4 は、まず、エミュレーション実行の準備として、解析対象のマクロを結合し、一繋がりマクロとする。これは、入力されたマクロが複数のマクロに分割されている場合があるためである。

[0034] 次に、実行部 1 5 4 は、エミュレーション実行のためのエントリーポイント（マクロの実行開始点）を探索する。実行部 1 5 4 は、エントリーポイントの探索として、まず、マクロ中を走査し、AutoOpenをはじめとする自動マクロを探索する。実行部 1 5 4 は、自動マクロが存在する場合には、この自動マクロをエントリーポイントとして抽出する。

[0035] そして、実行部 1 5 4 は、自動マクロが存在しない場合、マクロ中にイベントプロシージャが存在しないかを調べる。実行部 1 5 4 は、マクロ中にイベントプロシージャが存在する場合には、このイベントプロシージャをエントリーポイントとして抽出する。そして、実行部 1 5 4 は、抽出したエント

リーポイントから、エミュレータ 16（後述）を用いてマクロを実行することによって、実行トレースを取得する。そして、実行部 154 は、取得した実行トレースを実行トレース解析部 155 へ出力する。

[0036] 実行トレース解析部 155 は、実行部 154 から出力された実行トレースを解析し、特徴抽出を実施することによって、マクロ特徴量を取得する。実行トレース解析部 155 は、取得したマクロ特徴量をマクロ特徴量 DB 18 に格納する。したがって、実行部 154 及び実行トレース解析部 155 は、文書ファイル内のマクロからマクロの特徴量を取得する機能を有する。

[0037] テキスト抽出部 156 は、文書ファイルおよび文書ファイルから抽出された画像ファイルを入力として受け取り、テキストを抽出して出力する。まず、テキスト抽出部 156 は、文書ファイルを構文解析することによって、文書ファイル中に含まれるテキストを抽出する。この処理は、文書ファイルのファイルフォーマットの定義に従った構文解析によって実現できる。また、テキスト抽出部 156 は、画像ファイルに対して、既存の OCR (Optical Character Recognition: 光学文字認識) 技術を適用することにより、画像ファイル中のテキストを抽出する。この処理は、既存の OCR ライブラリや、Google (登録商標) Cloud Vision API などのサービスを用いて実現可能である。以降、この画像からテキスト抽出部 156 が抽出したテキストは、文書ファイルから抽出されたテキストと合わせて処理される。

[0038] テキスト分析部 157 は、文書ファイルから抽出されたテキストを入力として受け取り、自然言語処理によって特徴抽出を実施することで、テキスト特徴量を取得する。テキスト分析部 157 は、自然言語処理の処理として、分かち書き、不明語の除去、ストップワード除去、ステミング、同義語の統一、言語モデル抽出の処理を実行する。

[0039] テキスト分析部 157 は、まず、抽出されたテキスト中に含まれた文章を分かち書きする。分かち書きとは、文章を語ごとに分割することである。続いて、テキスト分析部 157 は、一定以上の長さの語を除去する。これは、悪性マクロ付き文書ファイルでは意味のないランダムな文字の羅列がしばし

ば用いられるためである。この際、テキスト分析部 157 は、長さを除去の基準とする代わりに、辞書を用いて、辞書に含まれていない語を除去するようにしてもよい。

[0040] そして、テキスト分析部 157 は、ストップワードの除去を行う。ストップワードとは、どの文書にも共通的に高頻度で出現する語のことで、精度の低下を招くため削除する。例えば、英語では、代名詞、助動詞、前置詞などが該当する。ストップワードの除去は、オープンデータのストップワードのコーパスを用いて、該当した語を除去することで実現できることが一般に知られている。

[0041] 続いて、テキスト分析部 157 は、ステミングを実施する。ステミングとは、語幹を抽出する処理であり、同じ語幹を持つが語形が変化している場合の処理である。例として、enabledやenablingは、ステミングによってenableに変換される。そして、テキスト分析部 157 は、同義語の統一処理を実施する。これは、同じ意味を持つが異なる単語が使われているものを統一し、より意味に焦点を当てるための処理である。例として、enableとpermitは一つの素性に統一される。これによって、単語の言い換えによる回避を防ぐことが期待できる。こうした同義語の統一は、同義語辞書を用いて実現できる。そして、テキスト分析部 157 は、これらの処理を施したテキストから、特徴抽出を実施する。したがって、テキスト抽出部 156 及びテキスト分析部 157 は、文書ファイル内のテキストからテキストの特徴量を取得する機能を有する。

[0042] なお、本実施の形態では、N-gramによる特徴抽出を実施した例について説明する。このN-gramとは、隣り合うN個の語を一要素としたモデルであり、要素の各々の出現回数を数え上げ、数え上げた出現回数を特徴量とする。テキスト分析部 157 は、上述した処理によって得られた特徴量をテキスト特徴量として、テキスト特徴量DB 19（後述）に格納する。なお、本実施の形態では、N-gramによる特徴抽出を実施した例について説明したが、もちろん、特徴抽出処理の設計はこれに限るものではない。

- [0043] クラスタ分析部 158 は、マクロ特徴量DB 18 及びテキスト特徴量DB 19 に格納された特徴量に基づいてクラスタリングを実施し、生成された各クラスタを出力する。クラスタ分析部 158 は、マクロ特徴量DB 18 及びテキスト特徴量DB 19 に格納されたデータに基づき、特徴ベクトルを生成する。
- [0044] この特徴ベクトルは、各データ点（すなわち、本実施の形態では、各マクロ付き文書ファイル）のそれぞれの特徴量をまとめてベクトル表現にしたものであり、すべてのデータ点の特徴量を共通した次元のベクトルで表せるようにしたものである。例えば、あるデータが、Aが「1」、Bが「2」、Cが「3」という特徴量を持ち、別のデータ点が、Cが「3」、Dが「4」、Eが「5」という特徴量を持つとき、これらを共通した次元のベクトルで表現するためには、（A, B, C, D, E）というベクトルである必要がある。これにより、前者のデータは（1, 2, 3, 0, 0）という特徴ベクトルになり、後者のデータは、（0, 0, 3, 4, 5）という特徴ベクトルになる。
- [0045] クラスタ分析部 158 は、マクロ特徴量とテキスト特徴量との双方に基づいたマルチモーダルクラスタリングを実現する必要がある。その実現方法として、2つの方法がある。第1の方法は、マクロ特徴量とテキスト特徴量とを事前に統合した特徴ベクトルを生成し、それに基づいてクラスタリングする方法である。第2の方法は、マクロ特徴量によるクラスタリングの結果とテキスト特徴量によるクラスタリングの結果とを統合する方法である。結果を統合する方法として、例えば、各結果のANDをとるものと、ORをとるもののが考えられるが、これに限るものではない。また、特徴量を統合する方法として、例えば、特徴ベクトルを連結する方法や、Deep AutoEncoderを用いる方法が考えられるが、これに限るものではない。
- [0046] クラスタリングアルゴリズムは、以下の要件に適合するものを選択する。第1に、多数の文書ファイルを高速に処理することを目的としているため、計算量が小さいアルゴリズムである必要がある。第2に、クラスタリング時にクラスタ数が判明していないため、クラスタ数を事前に設定する必要のな

いアルゴリズムを選択する必要がある。クラスタリング時にクラスタ数が判明しないのは、クラスタ数が、攻撃者がキャンペーン内で用いる悪性マクロ付き文書ファイルの種類数や、良性マクロ付き文書ファイルの種類数に左右されるためである。

[0047] そして、クラスタリングアルゴリズムについては、前述の2つの要件がある一方、階層構造は必要としない。したがって、クラスタリングアルゴリズムとして、計算量の要件から、階層型クラスタリングと比較して、一般に計算量が小さい非階層型クラスタリングのアルゴリズムが望まれる。また、クラスタ数がクラスタリング時に判明していないことから、事前にクラスタ数を与えるのではなく、予め設定した停止条件を満たすまでクラスタの分割を続けるアルゴリズムが適している。こうした条件を満たすクラスタリングアルゴリズムの一つにX-meansがあるが、これに限らず、K-meansを改造して独自の停止条件を付加したアルゴリズムを用いてもよく、その他のアルゴリズムを用いてもよい。

[0048] 選択部159は、クラスタ分析部158で生成されたクラスタを入力として受け取り、解析すべき文書ファイルを選定した結果を出力する。選択部159は、各クラスタの代表点のみを解析し、その解析による判定結果の悪性または良性をそのクラスタに属するデータ全体の判定とすることで、実質的に解析すべき対象を絞り込む。これは、マクロを内包した文書ファイル群をクラスタリングする際、悪性なものは強く凝縮されたクラスタを生成しやすく、かつ、良性なものと混合したクラスタになりづらいという特長による。

[0049] 選択部159は、各クラスタの重心を計算し、重心に最も近いデータ点をクラスタの代表点として用いる。このとき、距離計算には、ユークリッド距離を用いるのが一般的であるが、それ以外の距離尺度を用いてもよい。さらに、選択部159は、クラスタ内に属するデータ点の数を調べ、多いクラスタの代表点を優先的に詳細解析するように優先順位付けする。これは、キャンペーンにおいて、類似した悪性マクロ付き文書ファイルが大量に用いられるため、悪性なクラスタはクラスタ内のデータ点が多くなりやすいためであ

る。また、代表点の一ファイルの判定結果がクラスタ全体の判定結果となるため、より影響の大きいものから選ぶ意味もある。選択部 159 は、この処理によって得られた、優先順位付きの解析すべきファイルの一覧を、選択結果として出力する。

[0050] エミュレータ 16 は、マクロおよびエントリーポイントを受け取り、VBAの仕様にに基づき実行する機構である。エミュレータ 16 の詳細な構成は後述する。

[0051] ファイルハッシュDB 17 は、入力された文書ファイルに対応するハッシュ値を記憶するデータベースである。マクロ特徴量DB 18 は、実行トレース解析部 155 で抽出されたマクロ特徴量を記憶するデータベースである。テキスト特徴量DB 19 は、テキスト分析部 157 で抽出されたテキスト特徴量を記憶するデータベースである。これらのファイルハッシュDB 17、マクロ特徴量DB 18 及びテキスト特徴量DB 19 は、通信部 13 を介して制御部 15 と通信する、ストレージを内蔵する汎用のサーバ装置等であってもよい。

[0052] [ファイルハッシュDBが記憶するデータのデータ形式]

次に、ファイルハッシュDB 17 が記憶するデータのデータ形式の一例について説明する。図 2 は、図 1 に示すファイルハッシュDB 17 が記憶するデータのデータ形式の一例を示す図である。

[0053] 図 2 の表 T 1 に示すように、ファイルハッシュDB 17 は、スキーマとして、「id」、「file\_name」、「hash\_val」の 3 つの属性を持つ。「id」は、データベースに格納されているレコードの通し番号である。「file\_name」は、文書ファイルのファイル名である。「hash\_val」は、「file\_name」を持つファイルのハッシュ値である。

[0054] [マクロ特徴量DBが記憶するデータのデータ形式]

次に、マクロ特徴量DB 18 が記憶するデータのデータ形式の一例について説明する。図 3 は、図 1 に示すマクロ特徴量DB 18 が記憶するデータのデータ形式の一例を示す図である。なお、図 3 の表 T 2 に示す例では、特徴抽出に N-gram を用いている。ただし、N-gram 以外の方法によって特徴抽出が実行

されてもよい。

[0055] マクロ特微量DB 18は、スキーマとして、「id」、「hash\_val」、「ngram\_1」、「ngram\_...」、「ngram\_N」、「value」の属性を持つ。「id」は、データベースに格納されているレコードの通し番号である。「hash\_val」は、マクロが含まれていた文書ファイルのハッシュ値である。「ngram\_1」から「ngram\_N」までが、特徴を示す属性である。ここで、「N」には実際の数値が入り、「N-gram」の「N」の値に等しい。

[0056] 具体的には、3-gramによる特徴抽出を実施している場合、属性は「ngram\_1, ngram\_2, ngram\_3」となる。もちろん、4-gramによる特徴抽出であれば「ngram\_4」が追加され、5-gramであれば、さらに「ngram\_5」が追加される。また、「ngram\_1, ..., ngram\_X, ..., ngram\_N」において、「ngram\_X」は、N-gramのX番目の構成要素である。このため、3-gramでの特徴抽出において、「ngram\_2」は2番目の構成要素を示す。「value」は、対応する特微量を数値で格納する。N-gramによる特徴抽出では、対応するN-gramの要素が出現する回数となる。

[0057] 1レコードの具体例を挙げる。例えば、3-gramで特徴抽出するものとしたとき、あるマクロの実行トレース中に、(CreateObject, Adodb.Stream.open, Adodb.Stream.write)という順序の呼び出しが4回出現したとする。この場合には、ngram\_1が「CreateObject」、ngram\_2が「Adodb.Stream.open」、ngram\_3が「Adodb.Stream.write」となり、valueが「4」となる。なお、実行トレース解析部155がN-gram以外による特徴抽出をする場合は、ngram\_1からngram\_Nまでの部分を特徴抽出に応じた属性に変更する。

[0058] [テキスト特微量DBが記憶するデータのデータ形式]

次に、テキスト特微量DB 19が記憶するデータのデータ形式について説明する。図4は、図1に示すテキスト特微量DB 19が記憶するデータのデータ形式の一例を示す図である。図4の表T3に示す例では、特徴抽出にN-gramを用いている。ただし、N-gram以外の方法によって特徴抽出が実行されてもよい。

[0059] テキスト特徴量DB 19は、特徴抽出にN-gramを用いる場合には、スキーマとして、マクロ特徴量DB 18と同じく、「id」、「hash\_val」、「ngram\_1」、「ngram\_...」、「ngram\_N」、「value」の属性を持つ。また、N-gram以外による特徴抽出をする場合は、ngram\_1からngram\_Nまでの部分を特徴抽出に応じた属性に変更することも同様である。

[0060] [エミュレータの構成]

本実施の形態に係る選択装置10は、エミュレータ16を用いてマクロを実行することによって、実行トレースを取得する。選択装置10がエミュレータ16を必要とするのは、VBAの実行環境に、変更を加えられるサードパーティの実装が存在しないためである。また、柔軟な解析を実現するためには、エミュレータの実装が必要となる。

[0061] このため、図1に示すエミュレータ16の構成について説明する。図5は、図1に示すエミュレータ16の構成の一例を説明する図である。エミュレータ16は、字句解析部161、構文解析部162、中間表現生成部163、実行部164、APIハンドラ部165、オブジェクトハンドラ部166、ドキュメントインタフェース部167及び記録部168から構成される。

[0062] エミュレータ16は、入力としてVBAで記述されたマクロ（分析対象マクロ200）と、そのエントリーポイントとを受け取る。エミュレータ16は、受け取ったマクロを、エントリーポイントを起点としてエミュレーション実行し、実行トレースを出力する。

[0063] まず、字句解析部161は、マクロを走査してプログラムを構成する要素の最小単位（トークン）の並びとして認識する。

[0064] そして、構文解析部162は、字句解析結果を基に、VBAの仕様（詳細は、[MS-VBAL]: VBA Language Specification, Microsoft、[online]、[平成29年2月15日検索]、インターネット<URL: <https://msdn.microsoft.com/ja-jp/library/dd361851.aspx>>参照）に従い、マクロの構文解析を実施する。

[0065] 中間表現生成部163は、構文解析結果に基づいて、中間表現を生成する

。中間表現には、独自に定義した表現を用いてもよく、LLVM IRやVEX IRなどの既存の中間表現を用いてもよく、Pythonなど既存の言語で記述する形でもよい。

[0066] 実行部164は、生成された中間表現をエミュレーション実行していく。このエミュレーション実行の際には、変数をシンボルとして扱いながら実行する設計（シンボリック実行と呼ばれる（詳細は、V. Chipounov, V. Kuznetsov and G. Candea, “S2E: A Platform for In-Vivo Multi-Path Analysis of Software Systems”, In Proceedings of the Sixteenth International Conference on Architectural Support for Programming Languages and Operating Systems, pp. 265–278, 2011 参照））としてもよい。

[0067] ここで、シンボリック実行とは、プログラム実行時に変数をシンボルとして扱い、プログラムの実行結果を評価する手法である。シンボリック実行では、実行時には、シンボルを含んだ分岐条件を評価せず、プログラムの実行パスが網羅的に実行される。これにより、シンボリック実行では、解析妨害機能を具備したマクロを解析する場合であっても、妨害を回避し、網羅的な解析が期待できる。また、実行部164は、エミュレーション実行の際に、意図的に解析時間を長くするためのループが存在する場合に、検出して短縮するような機構を備えてもよい。

[0068] APIハンドラ部165は、エミュレーション実行中にAPI呼び出しがあった場合に、実行部164からAPI名及び引数を受け取り、それに対応した処理結果を返す。APIハンドラ部165は、例えば、Sleepなどを用いて意図的に解析時間を長くするような処理については、短縮するような処理としてもよい。また、APIハンドラ部165は、Eval関数が呼び出された場合、その引数であるマクロ（動的生成マクロ300）を、字句解析部161に入力する。Eval関数は、引数に与えられたマクロを実行する関数である。VBAは、動的言語であり、Eval関数によって動的生成されたコードが実行され得る。したがって、動的生成された動的生成マクロ300の追跡が必要となることから、API

ハンドラ部 165 は、この動的生成マクロ 300 を字句解析部 161 に入力する。

[0069] オブジェクトハンドラ部 166 は、マクロのエミュレーション実行中にActiveXオブジェクトを生成し利用した場合、オブジェクトに関連する情報を受け取り、対応した処理結果を返す。オブジェクトは、GetObject関数やCreateObject関数を用いて利用されるため、これらの呼び出しを基点とし、引数に基づいて動作する。オブジェクトに対応する処理の実現のために、ActiveXオブジェクトの種類ごとに対応する処理をエミュレータ 16 のオブジェクトハンドラ部 166 に予め定義しておく。或いは、オブジェクトに対応する処理の実現のために、エミュレータ 16 のオブジェクトハンドラ部 166 が、実際にActiveXオブジェクトを呼び出して仲介する方法を採用してもよい。

[0070] ドキュメントインタフェース部 167 は、マクロがエミュレーション実行中に文書ファイル 400 中の要素を参照した場合に、文書を構文解析して、参照された要素への操作を実施する。ドキュメントインタフェース部 167 は、例えば、Range関数やCell関数を用いて表中の特定のセルの値を取得するといった操作を実施する。悪性マクロ付き文書ファイルでは、文書内の特定の表中のセルにマクロの一部を保存しておき、マクロの実行時にこのセルを参照して動的に実行する、というような挙動が見られるため、このドキュメントインタフェース部 167 が必要となる。

[0071] 記録部 168 は、APIハンドラ部 165 とオブジェクトハンドラ部 166 とでのAPI呼び出しやオブジェクトメソッド呼び出しを監視し、実行トレースとして出力する。記録部 168 は、実行トレースとして、APIやオブジェクトメソッドの呼び出しと、その引数、返り値とを一組として時系列で記録する。

[0072] [選択装置の選択処理]

次に、選択装置 10 の選択処理の処理手順について説明する。図 6 は、図 1 に示す選択装置 10 の選択処理の処理手順を示すフローチャートである。

[0073] 図 6 に示すように、選択装置 10 は、まず、検査対象未知文書ファイル群 100 を入力として受け取る（ステップ S1）。そして、選択装置 10 は、

受け取った検査対象未知文書ファイル群100の中から、初めに処理するファイルを取り出す（ステップS2）。続いて、このファイルに対して、ハッシュ検査部151が、ファイルのハッシュ値を計算する（ステップS3）。そして、ハッシュ検査部151は、計算されたハッシュ値でファイルハッシュDB17を検索する（ステップS4）。

[0074] ハッシュ検査部151は、計算されたハッシュ値が検索に該当したか否かを判断する（ステップS5）。計算されたハッシュ値が検索に対して該当するとハッシュ検査部151が判断した場合（ステップS5：Yes）、制御部15は、そのファイルと同一のファイルが既に処理されており、これ以上の処理を必要としないものとして、次に処理するファイルを取り出す（ステップS6）。そして、制御部15は、ステップS3に戻り、このファイルに対するハッシュ値の計算処理を行う。

[0075] 一方、計算されたハッシュ値が検索に該当しないとハッシュ検査部151が判断した場合（ステップS5：No）、計算されたハッシュ値をファイルハッシュDB17に格納する（ステップS7）。次に、ファイル形式判定部152が、ヘッダやファイル構造を基にファイル形式を判定する（ステップS8）。そして、ファイル形式判定部152は、マクロを含まないファイル形式であるか否かを判断する（ステップS9）。ファイル形式判定部152は、マクロを含まないと判断した場合は（ステップS9：No）、選択装置10の処理対象外であるとして、次に処理するファイルを取り出し（ステップS6）、ステップS3に戻る。ファイル形式判定部152は、マクロを含むと判断した場合は（ステップS9：Yes）、要素抽出部153が、文書ファイルからマクロ及び画像を抽出する（ステップS10）。次に、テキスト抽出部156及びテキスト分析部157が、テキスト分析処理を実施する（ステップS11）。

[0076] 続いて、マクロ分析部15Aが、マクロ分析処理を実施する（ステップS12）。続いて、制御部15は、入力されたすべての文書ファイルを処理したか否かを判断する（ステップS13）。制御部15は、入力されたすべて

の文書ファイル进行处理していないと判断した場合（ステップS 1 3 : N o）、次に処理するファイルを取り出し（ステップS 6）、ステップS 3に戻る。

[0077] 一方、制御部 1 5 は、入力されたすべての文書ファイル进行处理したと制御部 1 5 が判断した場合（ステップS 1 3 : Y e s）、クラスタ分析部 1 5 8 は、クラスタ分析処理を実施する（ステップS 1 4）。そして、選択部 1 5 9 は、解析対象となる文書ファイルを選択する選択処理を実施し（ステップS 1 5）、得られた詳細解析対象の文書ファイルを出力して（ステップS 1 6）、終了する。

[0078] [テキスト分析処理の処理手順]

次に、テキスト分析処理の処理手順について説明する。このテキスト分析処理は、テキスト抽出部 1 5 6 が入力を受け取ってから、テキスト分析部 1 5 7 が処理を完了するまでの処理を含む。図 7 は、図 6 に示すテキスト分析処理の処理手順を示すフローチャートである。

[0079] まず、テキスト抽出部 1 5 6 は、文書ファイル及び該文書ファイルから抽出した画像ファイルを入力として受け取る（ステップS 2 1）。ここで、文書ファイル中に画像ファイルが存在する場合と存在しない場合がある。このため、テキスト抽出部 1 5 6 は、画像ファイルが入力として受け取ったファイル中に存在するか否かを判断する（ステップS 2 2）。

[0080] テキスト抽出部 1 5 6 は、画像ファイルが入力として受け取ったファイル中に存在すると判断した場合（ステップS 2 2 : Y e s）、OCRにより画像ファイルから、すべてのテキストを抽出する（ステップS 2 3）。テキスト抽出部 1 5 6 は、画像ファイルが文書ファイル中に存在しないと判断した場合（ステップS 2 2 : N o）、或いは、ステップS 2 3 終了後、文書ファイルに対する構文解析により、文書ファイルからすべてのテキストを抽出する（ステップS 2 4）。

[0081] そして、テキスト分析部 1 5 7 は、ステップS 2 3 およびステップS 2 4 の処理によって抽出されたテキスト群の中から、初めに処理するテキストを

取り出す（ステップS 2 5）。そして、テキスト分析部 1 5 7 は、そのテキストを分かち書きする（ステップS 2 6）。続いて、テキスト分析部 1 5 7 は、分かち書きによってできた語の並びから、先頭の語を取り出す（ステップS 2 7）。

[0082] テキスト分析部 1 5 7 は、取り出した語が不明語であるか否かを判断する（ステップS 2 8）。テキスト分析部 1 5 7 は、取り出した語が不明語であると判断した場合（ステップS 2 8 : Y e s）、この語は、除去対象であるため、処理を停止して、次の語を取り出し（ステップS 2 9）、ステップS 2 8に戻る。

[0083] 一方、テキスト分析部 1 5 7 は、取り出した語が不明語でないと判断した場合（ステップS 2 8 : N o）、この取り出した語がストップワードであるか否かを判断する（ステップS 3 0）。テキスト分析部 1 5 7 は、取り出した語がストップワードであると判断した場合（ステップS 3 0 : Y e s）、この語は、除去対象であるため、処理を停止して、次の語を取り出し（ステップS 2 9）、ステップS 2 8に戻る。

[0084] これに対し、テキスト分析部 1 5 7 は、取り出した語がストップワードでないと判断した場合（ステップS 3 0 : N o）、この取り出した語に対してステミングを実施する（ステップS 3 1）。さらに、テキスト分析部 1 5 7 は、この語に同義語が存在する場合には、同義語を統一する（ステップS 3 2）。そして、テキスト分析部 1 5 7 は、一テキスト中のすべての語を処理したか否かを判断する（ステップS 3 3）。

[0085] テキスト分析部 1 5 7 は、一テキスト中のすべての語を処理していないと判断した場合（ステップS 3 3 : N o）、次の語を取り出し（ステップS 2 9）、ステップS 2 8に戻る。これに対し、テキスト分析部 1 5 7 は、一テキスト中のすべての語を処理したと判断した場合（ステップS 3 3 : Y e s）、次にすべてのテキストを処理したかを判断する（ステップS 3 4）。

[0086] テキスト分析部 1 5 7 は、すべてのテキストを処理していないと判断した場合（ステップS 3 4 : N o）、次のテキストを取り出し（ステップS 3 5

）、ステップS 2 6に戻り、この取り出したテキストに対する処理を行う。一方、テキスト分析部1 5 7は、すべてのテキストを処理したと判断した場合（ステップS 3 4：Y e s）、N-gramにより特徴抽出を実施する（ステップS 3 6）。そして、テキスト分析部1 5 7は、抽出した特徴量をテキスト特徴量DB 1 9に格納し（ステップS 3 7）、テキスト分析処理を終了する。

[0087] [マクロ分析処理の処理手順]

次に、マクロ分析処理の処理手順について説明する。マクロ分析処理は、実行部1 5 4が入力を受け取ってから、実行トレース解析部1 5 5が処理を完了するまでの処理を含む。図8は、図6に示すマクロ分析処理の処理手順を示すフローチャートである。

[0088] まず、実行部1 5 4は、マクロを入力として受け取る（ステップS 4 1）。そして、実行部1 5 4は、マクロが複数存在するか否かを判断する（ステップS 4 2）。実行部1 5 4は、マクロが複数存在すると判断した場合（ステップS 4 2：Y e s）、これらの複数のマクロを結合する（ステップS 4 3）。実行部1 5 4は、ステップS 4 3の処理終了後、或いは、マクロが複数存在しないと判断した場合（ステップS 4 2：N o）、マクロを走査し、エントリーポイントを探索する（ステップS 4 4）。

[0089] そして、実行部1 5 4は、エミュレータ1 6に、エミュレーション実行処理を実施させる（ステップS 4 5）。そして、実行部1 5 4は、エミュレーション実行処理によって得られた実行トレースからAPI/メソッド呼び出しの列を受け取る（ステップS 4 6）。実行部1 5 4は、受け取ったAPI/メソッド呼び出しの列を用いて、N-gramにより特徴抽出を行う（ステップS 4 7）。そして、実行部1 5 4は、抽出した特徴量をマクロ特徴量DB 1 8に格納し（ステップS 4 8）、マクロ分析処理を終了する。

[0090] [エミュレーション実行処理の流れ]

次に、図8に示すエミュレーション実行処理の処理手順について説明する。図9は、図8に示すエミュレータ実行処理の処理手順を示すフローチャートである。

- [0091] まず、エミュレータ 16 は、マクロおよびエン트리ポイントを入力として受け取る（ステップ S 5 1）。そして、エミュレータ 16 では、字句解析部 161 がマクロを字句解析する（ステップ S 5 2）。さらに、構文解析部 162 は、字句解析の結果に対して、構文解析を実施する（ステップ S 5 3）。続いて、中間表現生成部 163 は、字句解析の結果及び構文解析の結果に基づいて、中間表現を生成する（ステップ S 5 4）。そして、エミュレータ 16 は、中間表現の初めのステップを取り出す（ステップ S 5 5）。ここで、初めのステップとは、中間表現中でエン트리ポイントに対応する部分である。実行部 164 は、この中間表現のステップの処理を実行する（ステップ S 5 6）。
- [0092] ここで、エミュレータ 16 は、API 呼び出しがあったか否かを判断する（ステップ S 5 7）。API 呼び出しがあったとエミュレータ 16 が判断した場合（ステップ S 5 7 : Yes）、API ハンドラ部 165 が API に対応した処理を実行し（ステップ S 5 8）、記録部 168 が API 呼び出しの情報を記録する（ステップ S 5 9）。
- [0093] そして、エミュレータ 16 は、ステップ S 5 9 終了後、或いは、API 呼び出しがないと判断した場合（ステップ S 5 7 : No）、Eval 関数が呼び出されたか否かを判断する（ステップ S 6 0）。Eval 関数が呼び出されたとエミュレータ 16 が判断した場合（ステップ S 6 0 : Yes）、この引数を動的に生成された新たなマクロとし（ステップ S 6 1）、実行処理に移る。具体的には、この新たなマクロの先頭をエン트리ポイントとして（ステップ S 6 2）、実行部 164 が、ステップ S 5 2～ステップ S 7 0 の処理を順次実行するエミュレーション実行処理を行い（ステップ S 6 3）、中間表現の次のステップを取り出した後（ステップ S 6 4）、ステップ S 5 6 に進み、以降の処理を実行する。
- [0094] 一方、エミュレータ 16 は、Eval 関数が呼び出されていないと判断した場合（ステップ S 6 0 : No）、ActiveX オブジェクトの呼び出しがあるか判断する（ステップ S 6 5）。ActiveX オブジェクト呼び出しがあるとエミュレー

タ 1 6 が判断した場合（ステップ S 6 5 : Y e s）、オブジェクトハンドラ部 1 6 6 がオブジェクトに対応した処理を実行し（ステップ S 6 6）、記録部 1 6 8 がオブジェクトの生成やオブジェクトメソッドの呼び出しに関連する情報を記録する（ステップ S 6 7）。

[0095] エミュレータ 1 6 は、ActiveX オブジェクト呼び出しがないと判断した場合（ステップ S 6 5 : N o）、或いは、ステップ S 6 7 の処理終了後、文書ファイルが参照されたかを判断する（ステップ S 6 8）。エミュレータ 1 6 は、文書ファイルが参照されたと判断した場合（ステップ S 6 8 : Y e s）、ドキュメントインタフェース部 1 6 7 が文書への処理、すなわち、文書を構文解析して、参照された要素への操作を実施する（ステップ S 6 9）。

[0096] エミュレータ 1 6 は、文書ファイルが参照されていないと判断した場合（ステップ S 6 8 : N o）、或いは、ステップ S 6 9 の処理終了後、中間表現をすべて実行したか否かを判断する（ステップ S 7 0）。エミュレータ 1 6 は、まだ中間表現をすべて実行していないと判断した場合（ステップ S 7 0 : N o）、中間表現の次のステップを取り出し（ステップ S 6 4）、ステップ S 5 6 に進み、以降の処理を実行する。一方、エミュレータ 1 6 は、まだ中間表現をすべて実行したと判断した場合（ステップ S 7 0 : Y e s）、実行トレースを実行トレース解析部 1 5 5 に出力し（ステップ S 7 1）、エミュレーション実行処理を終了する。

[0097] [クラスタ分析処理の処理手順]

次に、クラスタ分析処理の処理手順の一例について説明する。図 1 0 は、図 6 に示すクラスタ分析処理の処理手順を示すフローチャートである。図 1 0 は、選択装置 1 0 が、マクロ特徴量とテキスト特徴量とを事前に統合してクラスタ分析する場合のクラスタ分析処理の処理手順を示す。

[0098] まず、クラスタ分析部 1 5 8 は、マクロ特徴量 DB 1 8 から全特徴量を取り出す（ステップ S 8 1）。そして、クラスタ分析部 1 5 8 は、1 ファイルあたり 1 つのマクロに関する特徴ベクトルを持つように、1 ファイルごとに 1 つのマクロ特徴ベクトルを生成する（ステップ S 8 2）。続いて、クラスタ

分析部 158 は、テキスト特徴量 DB 19 から全特徴量を取り出し（ステップ S 83）、1 ファイルごとに 1 つのテキスト特徴ベクトルを生成する（ステップ S 84）。

[0099] そして、クラスタ分析部 158 は、ファイルごとにマクロ特徴ベクトルとテキスト特徴ベクトルとを統合し（ステップ S 85）、クラスタリングを実施する（ステップ S 86）。そして、クラスタ分析部 158 は、生成されたクラスタを出力し（ステップ S 87）、クラスタ分析処理を終了する。

[0100] また、クラスタ分析処理の処理手順の他の例について説明する。図 11 は、図 6 に示すクラスタ分析処理の他の処理手順を示すフローチャートである。図 11 では、選択装置 10 が、マクロ特徴量とテキスト特徴量でそれぞれクラスタリングを実施し、各々の結果を統合する場合のクラスタ分析処理の処理手順を示す。

[0101] まず、図 11 に示すステップ S 91～ステップ S 94 は、図 10 に示すステップ S 81～ステップ S 84 である。クラスタ分析部 158 は、マクロ特徴ベクトルでクラスタリングを実施する（ステップ S 95）。そして、クラスタ分析部 158 は、テキスト特徴ベクトルでクラスタリングを実施する（ステップ S 96）。クラスタ分析部 158 は、生成された各々のクラスタリングの結果を統合し（ステップ S 97）、クラスタを出力して（ステップ S 98）、クラスタ分析処理を終了する。

[0102] [選択処理の処理手順]

次に、図 6 に示す選択処理の処理手順について説明する。図 12 は、図 6 に示す選択処理の処理手順を示すフローチャートである。この選択処理において、選択部 159 は、クラスタ分析部 158 が生成したクラスタを入力として受け取り、解析すべき文書ファイルを選定した結果を出力する。

[0103] まず、選択部 159 は、クラスタ分析部 158 から、クラスタを入力として受け取る（ステップ S 101）。そして、選択部 159 は、初めに処理するクラスタを取り出す（ステップ S 102）。そして、選択部 159 は、クラスタの重心を計算し（ステップ S 103）、その重心に最も近いデータ点

をクラスタの代表点として抽出する（ステップS 1 0 4）。続いて、選択部 1 5 9 は、クラスタ内のデータ点の数を計測する（ステップS 1 0 5）。選択部 1 5 9 は、すべてのクラスタを処理しているか否かを判断する（ステップS 1 0 6）。

[0104] 選択部 1 5 9 は、すべてのクラスタを処理していないと判断した場合（ステップS 1 0 6 : N o）、次に処理するクラスタを取り出し（ステップS 1 0 7）、ステップS 1 0 3に進む。一方、選択部 1 5 9 は、すべてのクラスタを処理したと判断した場合（ステップS 1 0 6 : Y e s）、抽出した各クラスタの代表点をリスト化し（ステップS 1 0 8）、クラスタ内のデータ点の数で各代表点に優先順位付けする（ステップS 1 0 9）。そして、選択部 1 5 9 は、それを、詳細解析対象の文書ファイルに優先順位付けしたものとして出力し（ステップS 1 1 0）、選択処理を終了する。

[0105] [実施の形態の効果]

以上のように、本実施の形態では、文書ファイル内のマクロからマクロの特徴量を取得し、文書ファイル内のテキストからテキストの特徴量を取得し、取得したマクロの特徴量とテキストの特徴量とを用いてクラスタリングを実施して、解析対象の文書ファイルを絞り込んでいる。すなわち、本実施の形態では、マクロのエミュレーション実行に基づく特徴抽出と、文書ファイル中に含まれるテキストの分析に基づく特徴抽出とを行っている。そして、本実施の形態では、抽出されたマクロ特徴量とテキスト特徴量との双方に基づき、マルチモーダルクラスタリングを実施し、生成されたクラスタに基づいて解析をするべき対象となる文書ファイルを選別している。

[0106] したがって、本実施の形態では、エミュレーション実行により、難読化の妨害を受けることなく、攻撃者の用いる悪性マクロ付き文書ファイルの特徴を捉えることができる。そして、本実施の形態では、中期的観測でなく、マクロ部分のみの実行による解析により、中期的観測と比較して高速な特徴抽出が可能である。また、本実施の形態では、テキスト分析についても、軽量な特徴抽出手法により、特徴抽出の高速化を期待できる。

- [0107] このように、本実施の形態によれば、高速な特徴抽出によって類似検体を同一クラスタに分類し、解析する対象をクラスタの代表点のみに絞り込むことができるため、解析のプロセス全体の効率化を実現できる。さらに、本実施の形態によれば、マクロの挙動及びテキストという、攻撃に必須の特徴が現れる要素を網羅的に捉えた特徴抽出に基づくマルチモーダルクラスタリングにより、高精度な分類を実現できる。
- [0108] 上述したように、本実施の形態は、攻撃の特徴が現れる、テキストの語とマクロの挙動との双方の特徴を反映した分類に有用であり、多数の文書ファイルに対して解析すべき対象を高速に選定することに適している。このため、本実施の形態に係る選択装置10を、サンドボックスと共に組織内ネットワークの入口に導入することによって、ネットワークを通過する多数の文書ファイルを効率的に解析することが可能である。
- [0109] また、本実施の形態においては、おもにVBAで記述されたマクロを内包しているMicrosoft Office文書ファイルを前提としているが、もちろん、これに限らない。本実施の形態に係る選択装置10は、他の形式の文書ファイルである場合や、VBA以外で記述されたプログラムである場合であっても、テキストとプログラムの両方を含む形態の文書ファイルであれば、同様に選択可能である。これは、テキストとプログラムの両方が存在すれば、テキストの特徴量とプログラムの特徴量を抽出して、同様の手順でクラスタ分析し、代表点を抽出することが可能であることによる。したがって、本実施の形態に係る選択装置10は、例えば、悪性JavaScript（登録商標）を含むPDF形式のファイルの選択にも応用することが可能である。この場合、悪性マクロの代わりに悪性JavaScriptを分析し、MS Office文書ファイル中のテキストの代わりにPDFファイル中のテキストを分析すればよい。また、本実施の形態においては、エミュレータを用いて実行トレースを取得しているが、もちろん、エミュレータを用いずに他の方法で実行トレースを取得してもよい。例えば、Microsoft OfficeにAPI呼び出しやActiveXオブジェクトの生成およびメソッド呼び出しなどを監視して記録する変更を加え、実行トレースを取得しても

よい。また、たとえば、実行トレース取得機能を持った仮想マシンモニタを用いて、監視されたゲストOS (Operating System) 上のMS Officeで悪性マクロ付き文書ファイルを開くことで実行トレースを取得してもよい。また、実行トレース取得機能を、MS OfficeへのAPIフックや動的バイナリ計装 (Dynamic Binary Instrumentation) などを実現してもよい。

[0110] [他の実施の形態]

[入れ子構造の文書ファイルへの対応]

本実施形態では単一の文書ファイルのみを対象としたが、入れ子構造となっている文書ファイルを対象としてもよい。入れ子構造となっている文書ファイルとは、たとえば、PDFファイルの中に悪性マクロ付き文書ファイルが埋め込まれたものである。この場合、PDFファイルを開くと、PDFファイルの持つJavaScriptが実行され、それによって埋め込まれた悪性マクロ付き文書ファイルが保存され開かれることで、前述の攻撃が開始される。このような場合、選択装置に、埋め込まれたファイルを取り出す機能を追加することで対応できる。前述の例では、JavaScriptを実行して埋め込まれた悪性マクロ付き文書ファイルを抽出する機能を追加する。または、ファイルのシグネチャやヘッダ、フッタの情報に基づいて抽出する（ファイルカービングと呼ばれる）手法を用いてもよい。上述の機能によってあらかじめ悪性マクロ付き文書ファイルを取り出し、実施例1に記載の手法で解析することで、入れ子構造の文書ファイルに対応できる。

[0111] [システム構成等]

図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散・統合の具体的形態は図示のものに限られず、その全部又は一部を、各種の負荷や使用状況等に応じて、任意の単位で機能的又は物理的に分散・統合して構成することができる。さらに、各装置にて行なわれる各処理機能は、その全部又は任意の一部が、CPU及び当該CPUにて解析実行されるプログラムにて実現され、或いは、ワイヤードロジックによるハードウェアとして実現され得

る。

[0112] また、本実施形態において説明した各処理のうち、自動的に行われるものとして説明した処理の全部又は一部を手動的におこなうこともでき、或いは、手動的におこなわれるものとして説明した処理の全部又は一部を公知の方法で自動的におこなうこともできる。この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。

[0113] [プログラム]

図13は、プログラムが実行されることにより、選択装置10が実現されるコンピュータの一例を示す図である。コンピュータ1000は、例えば、メモリ1010、CPU1020を有する。また、コンピュータ1000は、ハードディスクドライバインタフェース1030、ディスクドライバインタフェース1040、シリアルポートインタフェース1050、ビデオアダプタ1060、ネットワークインタフェース1070を有する。これらの各部は、バス1080によって接続される。

[0114] メモリ1010は、ROM (Read Only Memory) 1011及びRAM1012を含む。ROM1011は、例えば、BIOS (Basic Input Output System) 等のブートプログラムを記憶する。ハードディスクドライバインタフェース1030は、ハードディスクドライブ1090に接続される。ディスクドライバインタフェース1040は、ディスクドライブ1100に接続される。例えば磁気ディスクや光ディスク等の着脱可能な記憶媒体が、ディスクドライブ1100に挿入される。シリアルポートインタフェース1050は、例えばマウス1110、キーボード1120に接続される。ビデオアダプタ1060は、例えばディスプレイ1130に接続される。

[0115] ハードディスクドライブ1090は、例えば、OS1091、アプリケーションプログラム1092、プログラムモジュール1093、プログラムデータ1094を記憶する。すなわち、選択装置10の各処理を規定するプログラムは、コンピュータ1000により実行可能なコードが記述されたプログ

ラムモジュール１０９３として実装される。プログラムモジュール１０９３は、例えばハードディスクドライブ１０９０に記憶される。例えば、選択装置１０における機能構成と同様の処理を実行するためのプログラムモジュール１０９３が、ハードディスクドライブ１０９０に記憶される。なお、ハードディスクドライブ１０９０は、SSDにより代替されてもよい。

[0116] また、上述した実施形態の処理で用いられる設定データは、プログラムデータ１０９４として、例えばメモリ１０１０やハードディスクドライブ１０９０に記憶される。そして、CPU１０２０が、メモリ１０１０やハードディスクドライブ１０９０に記憶されたプログラムモジュール１０９３やプログラムデータ１０９４を必要に応じてRAM１０１２に読み出して実行する。

[0117] なお、プログラムモジュール１０９３やプログラムデータ１０９４は、ハードディスクドライブ１０９０に記憶される場合に限らず、例えば着脱可能な記憶媒体に記憶され、ディスクドライブ１１００等を介してCPU１０２０によって読み出されてもよい。或いは、プログラムモジュール１０９３及びプログラムデータ１０９４は、ネットワーク（LAN、WAN等）を介して接続された他のコンピュータに記憶されてもよい。そして、プログラムモジュール１０９３及びプログラムデータ１０９４は、他のコンピュータから、ネットワークインタフェース１０７０を介してCPU１０２０によって読み出されてもよい。

[0118] 以上、本発明者によってなされた発明を適用した実施形態について説明したが、本実施形態による本発明の開示の一部をなす記述及び図面により本発明は限定されることはない。すなわち、本実施形態に基づいて当業者等によりなされる他の実施形態、実施例及び運用技術等はすべて本発明の範疇に含まれる。

## 符号の説明

- [0119]    １０   選択装置  
          １１   入力部  
          １２   出力部

- 1 3 通信部
- 1 4 記憶部
- 1 5 制御部
- 1 5 A マクロ分析部
- 1 6 エミュレータ
- 1 7 ファイルハッシュデータベース (DB)
- 1 8 マクロ特徴量DB
- 1 9 テキスト特徴量DB
- 1 5 1 ハッシュ検査部
- 1 5 2 ファイル形式判定部
- 1 5 3 要素抽出部
- 1 5 4、1 6 4 実行部
- 1 5 5 実行トレース解析部
- 1 5 6 テキスト抽出部
- 1 5 7 テキスト分析部
- 1 5 8 クラスタ分析部
- 1 5 9 選択部
- 1 6 1 字句解析部
- 1 6 2 構文解析部
- 1 6 3 中間表現生成部
- 1 6 5 APIハンドラ部
- 1 6 6 オブジェクトハンドラ部
- 1 6 7 ドキュメントインタフェース部
- 1 6 8 記録部

## 請求の範囲

- [請求項1]           マクロ付き文書ファイル内のマクロからマクロの特徴量を取得するマクロ分析部と、
- 前記文書ファイル内のテキストからテキストの特徴量を取得するテキスト分析部と、
- 前記マクロの特徴量と前記テキストの特徴量とを用いてクラスタリングを実施するクラスタ分析部と、
- 前記クラスタリングの結果に基づいて、解析対象の文書ファイルを選択する選択部と、
- を有することを特徴とする選択装置。
- [請求項2]           前記マクロ分析部は、前記マクロの実行により取得されたマクロのプロファイル情報である実行トレースの解析により前記マクロの特徴量を取得することを特徴とする請求項1に記載の選択装置。
- [請求項3]           前記マクロ分析部は、前記マクロの実行にエミュレータを用いることを特徴とする請求項1または2に記載の選択装置。
- [請求項4]           前記テキスト分析部は、前記文書ファイルに含まれる語に対する自然言語処理に基づいて前記テキストの特徴量を取得することを特徴とする請求項1～3のいずれか一つに記載の選択装置。
- [請求項5]           前記クラスタ分析部は、前記マクロの特徴量と前記テキストの特徴量とに基づいたマルチモーダルクラスタリングを実施することを特徴とする請求項1～4のいずれか一つに記載の選択装置。
- [請求項6]           前記マクロ分析部は、実行トレースとしてAPI (Application Programming Interface) 呼び出しおよびオブジェクトメソッド呼び出しを取得し、
- 前記テキスト分析部は、前記文書ファイルに含まれる語の種類と前記文書ファイルに含まれる語の並びとの特徴を抽出することを特徴とする請求項2～5のいずれか一つに記載の選択装置。
- [請求項7]           選択装置が実行する選択方法であって、

マクロ付き文書ファイル内のマクロからマクロの特徴量を取得するマクロ分析工程と、

前記文書ファイル内のテキストからテキストの特徴量を取得するテキスト分析工程と、

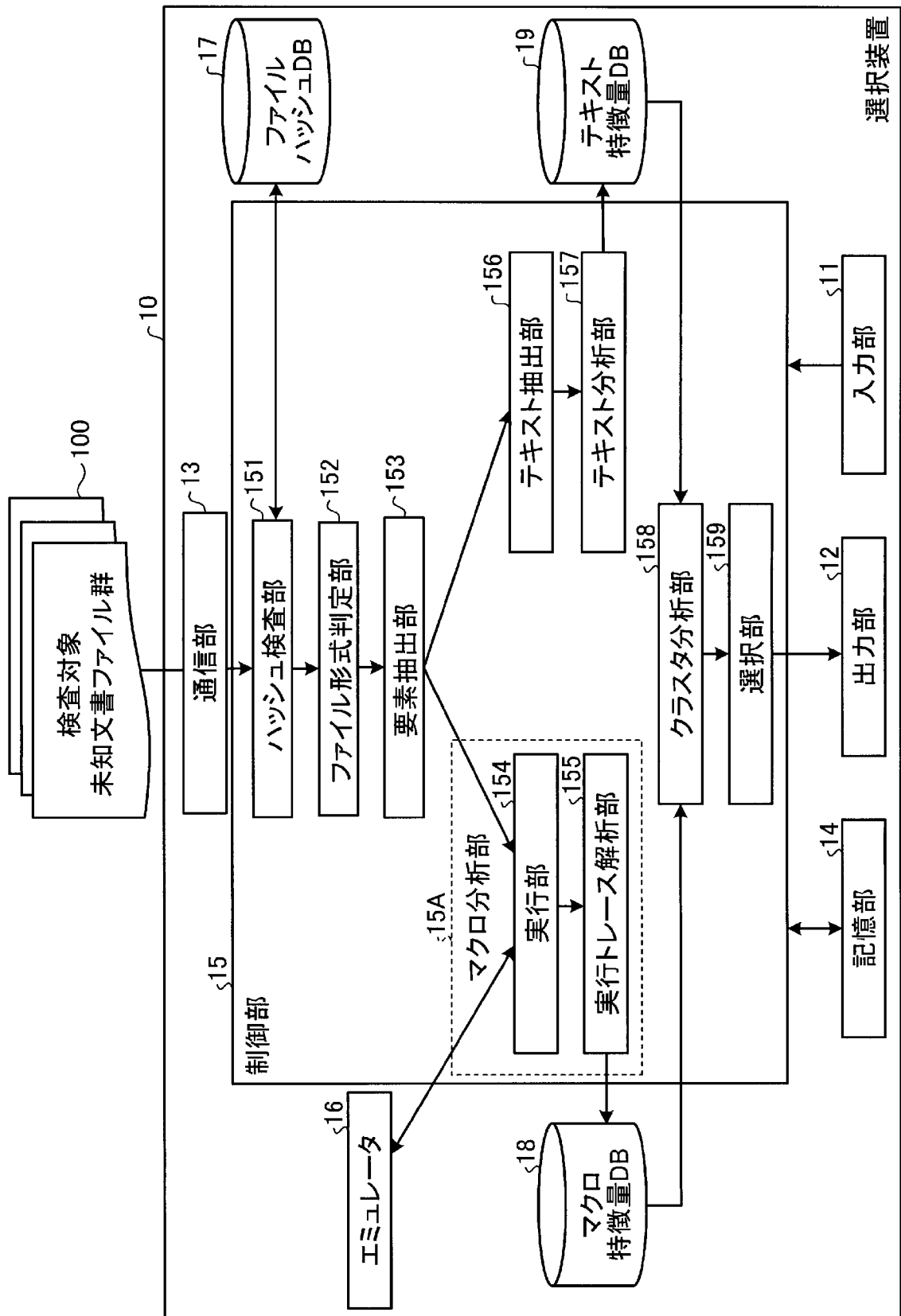
前記マクロの特徴量と前記テキストの特徴量とを用いてクラスタリングを実施するクラスタ工程と、

クラスタリングの結果に基づいて、解析対象の文書ファイルを選択する選択工程と、

を含んだことを特徴とする選択方法。

[請求項8]        コンピュータを、請求項1～6のいずれか一つに記載の選択装置として機能させるための選択プログラム。

[図1]



[図2]

T1

| id  | file_name   | hash_val    |
|-----|-------------|-------------|
| 1   | Invoice.doc | 5055803a... |
| ... | ...         | ...         |

[図3]

T2

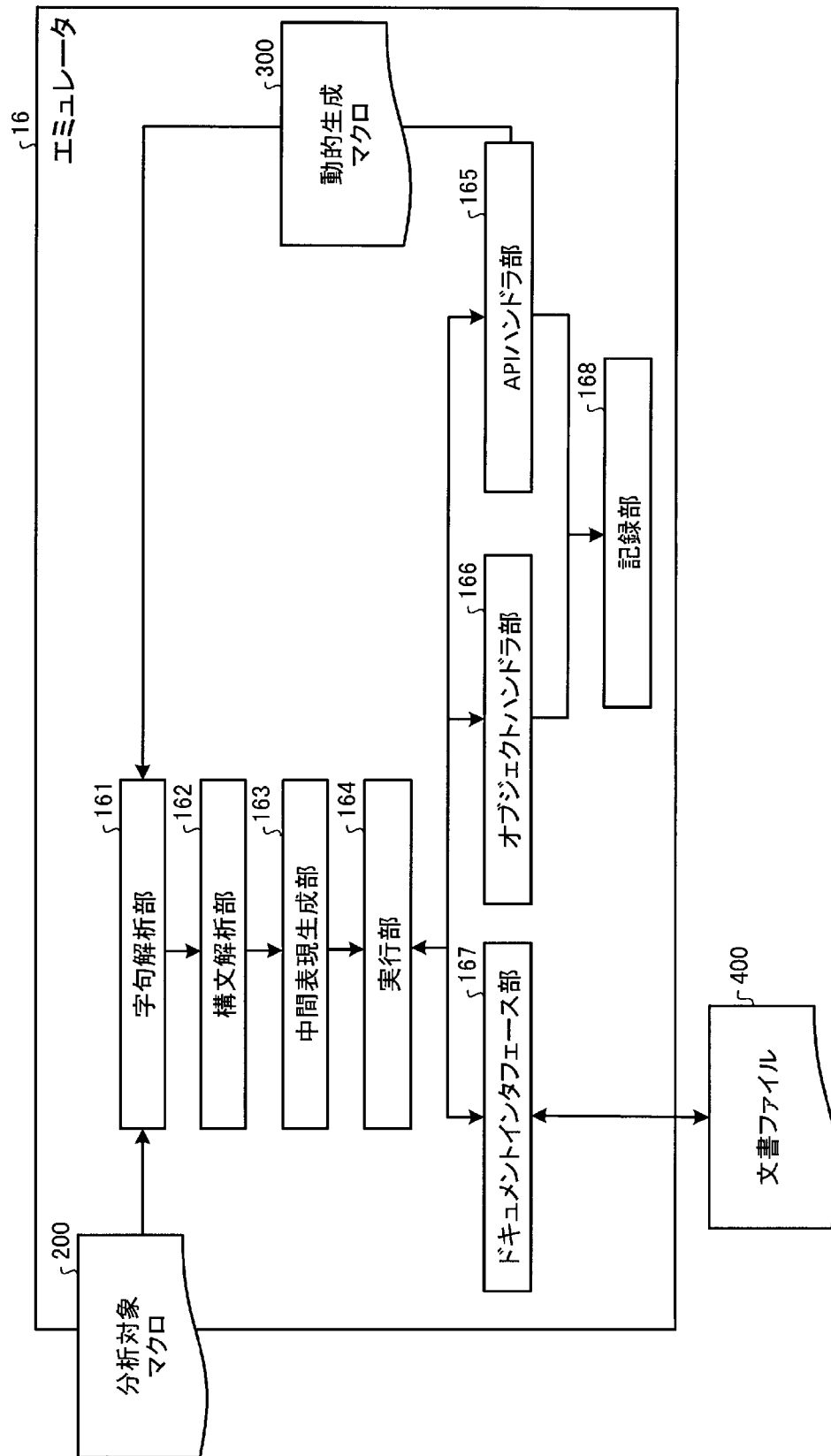
| id  | hash_val    | ngram_1           | ngram_... | ngram_N            | value |
|-----|-------------|-------------------|-----------|--------------------|-------|
| 1   | 5055803a... | CreateObject      | ...       | Adodb.Stream.write | 1     |
| 2   | 5055803a... | Adodb.Stream.Open | ...       | CreateObject       | 1     |
| ... | ...         | ...               | ...       | ...                | ...   |

[図4]

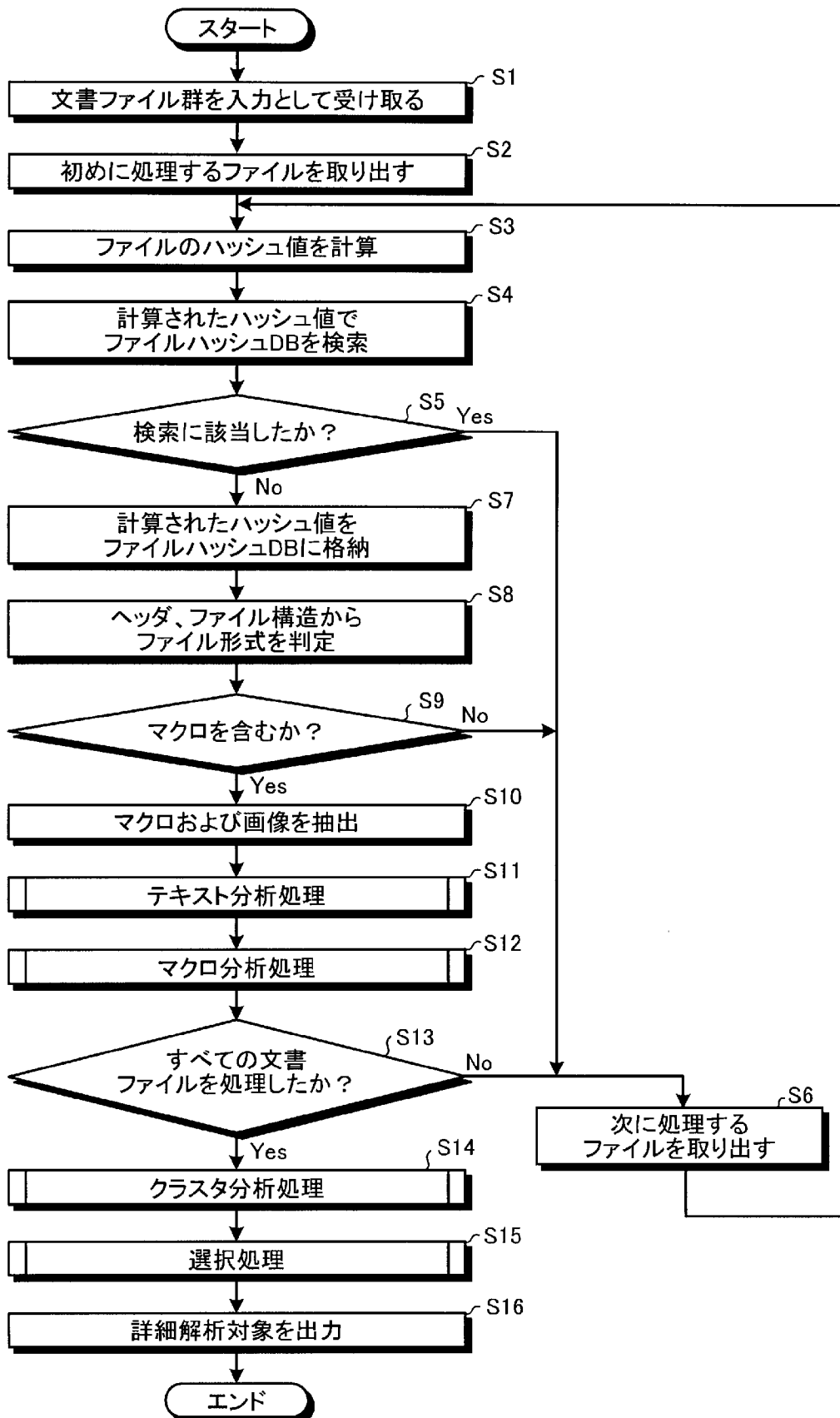
T3

| id  | hash_val    | ngram_1 | ngram_... | ngram_N  | value |
|-----|-------------|---------|-----------|----------|-------|
| 1   | 5055803a... | pleas   | ...       | document | 1     |
| 2   | 5055803a... | enabl   | ...       | view     | 1     |
| ... | ...         | ...     | ...       | ...      | ...   |

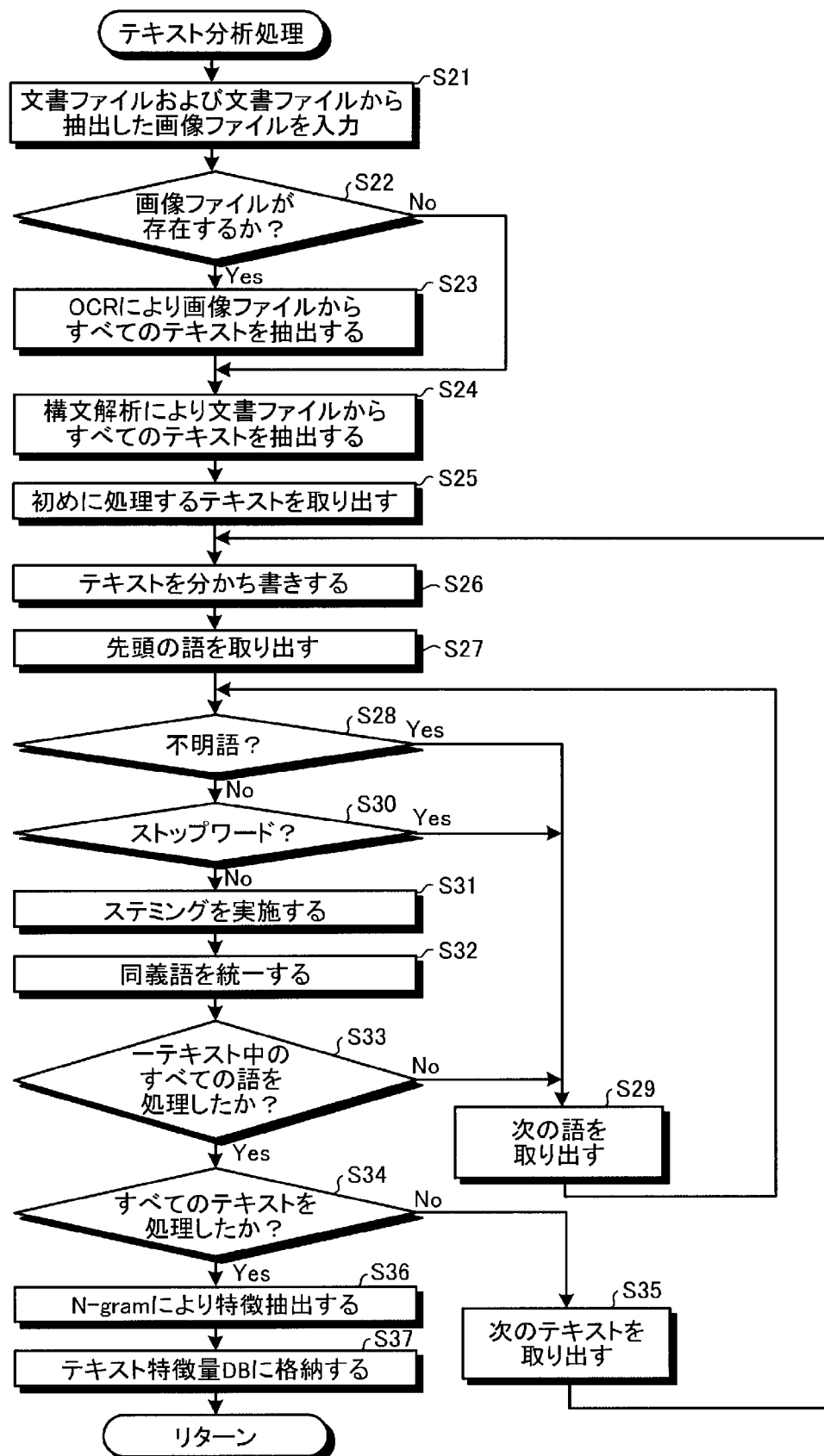
[図5]



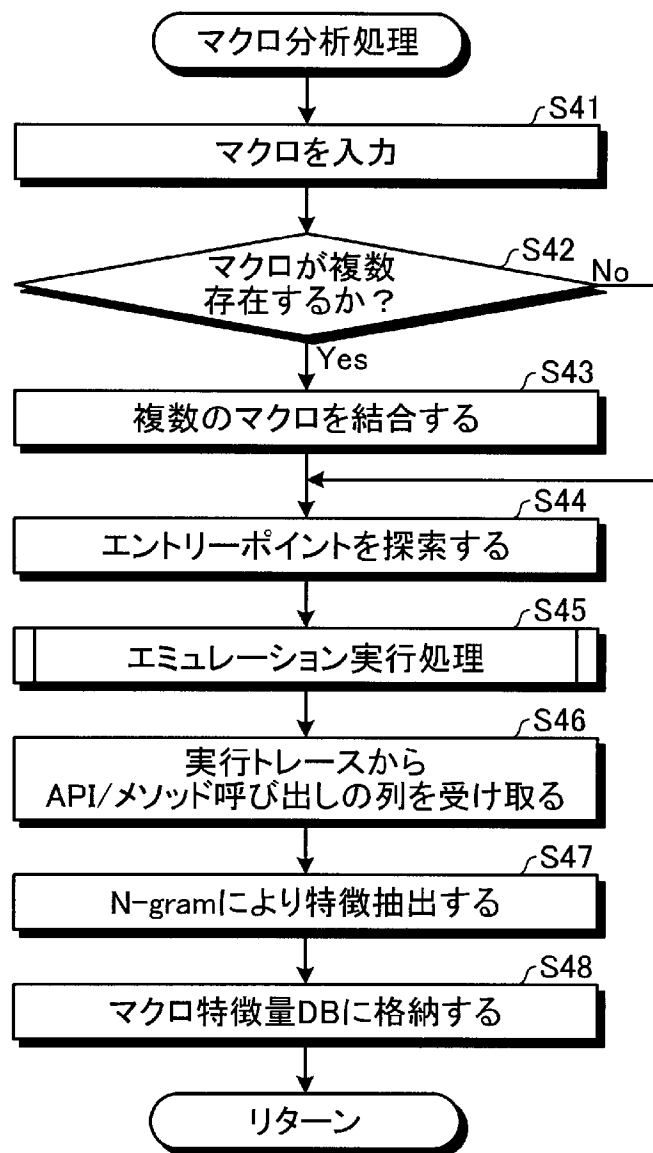
[図6]



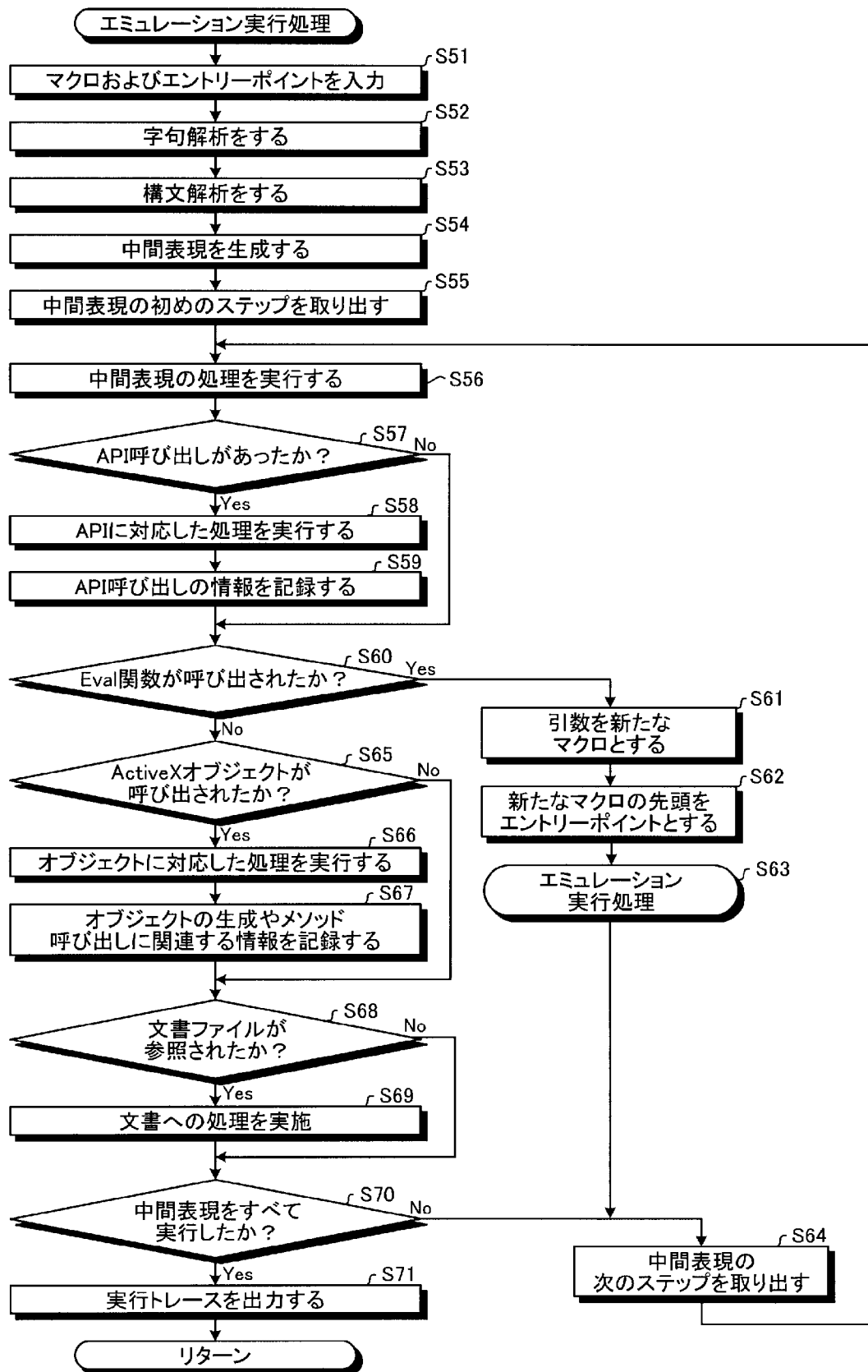
[図7]



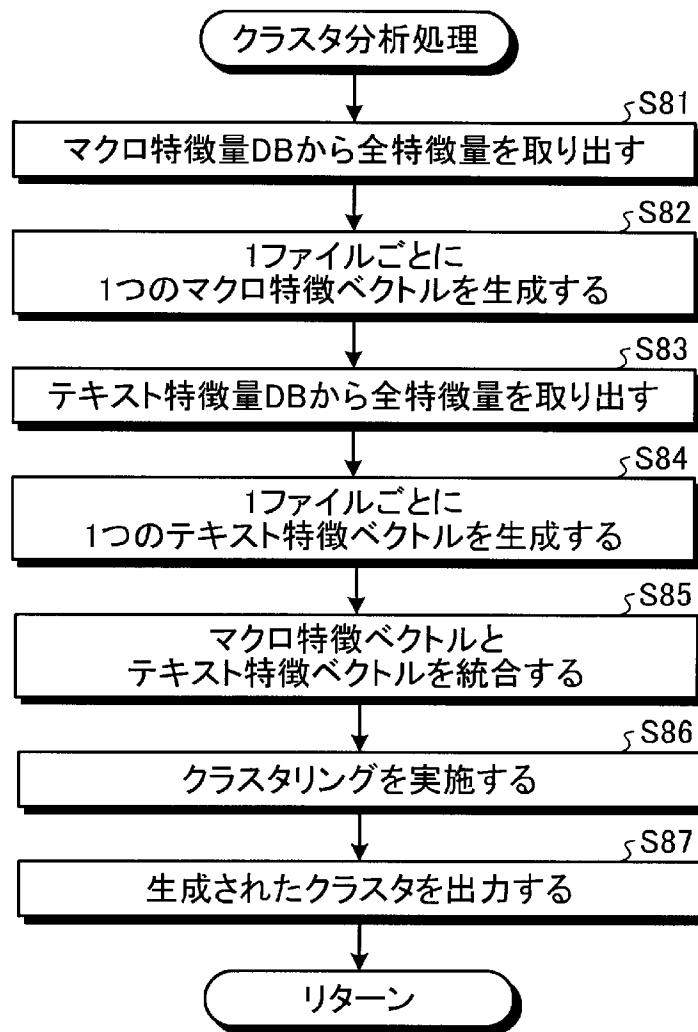
[図8]



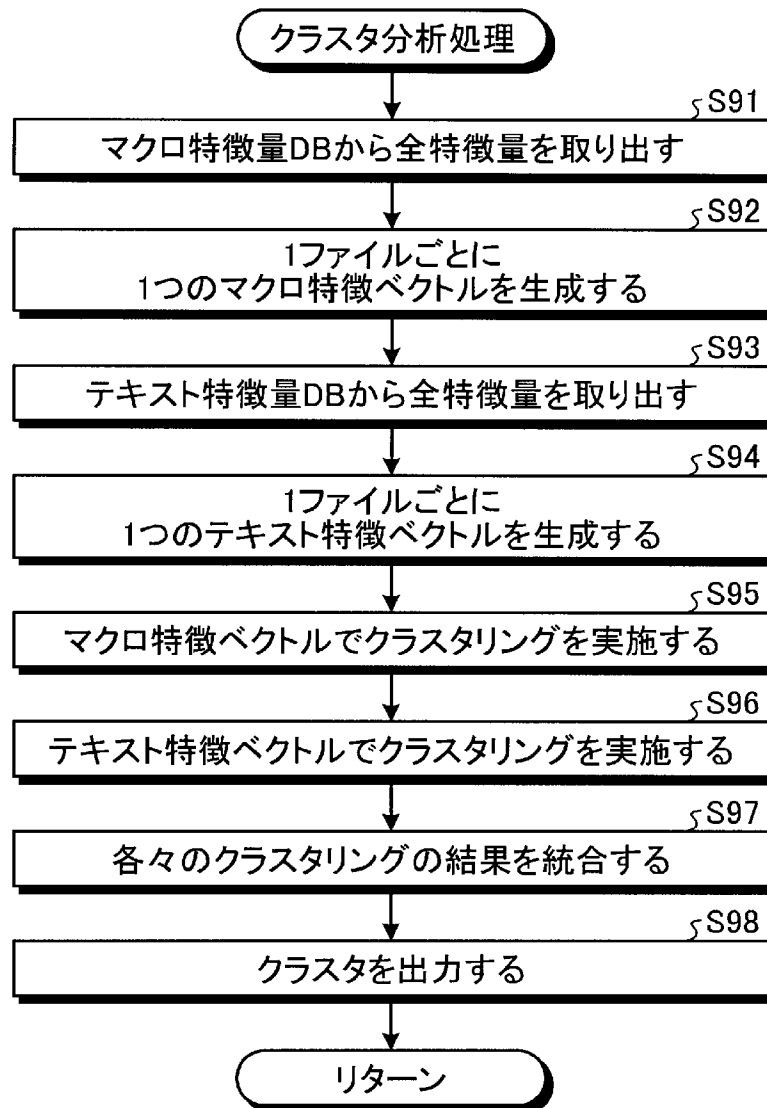
[図9]



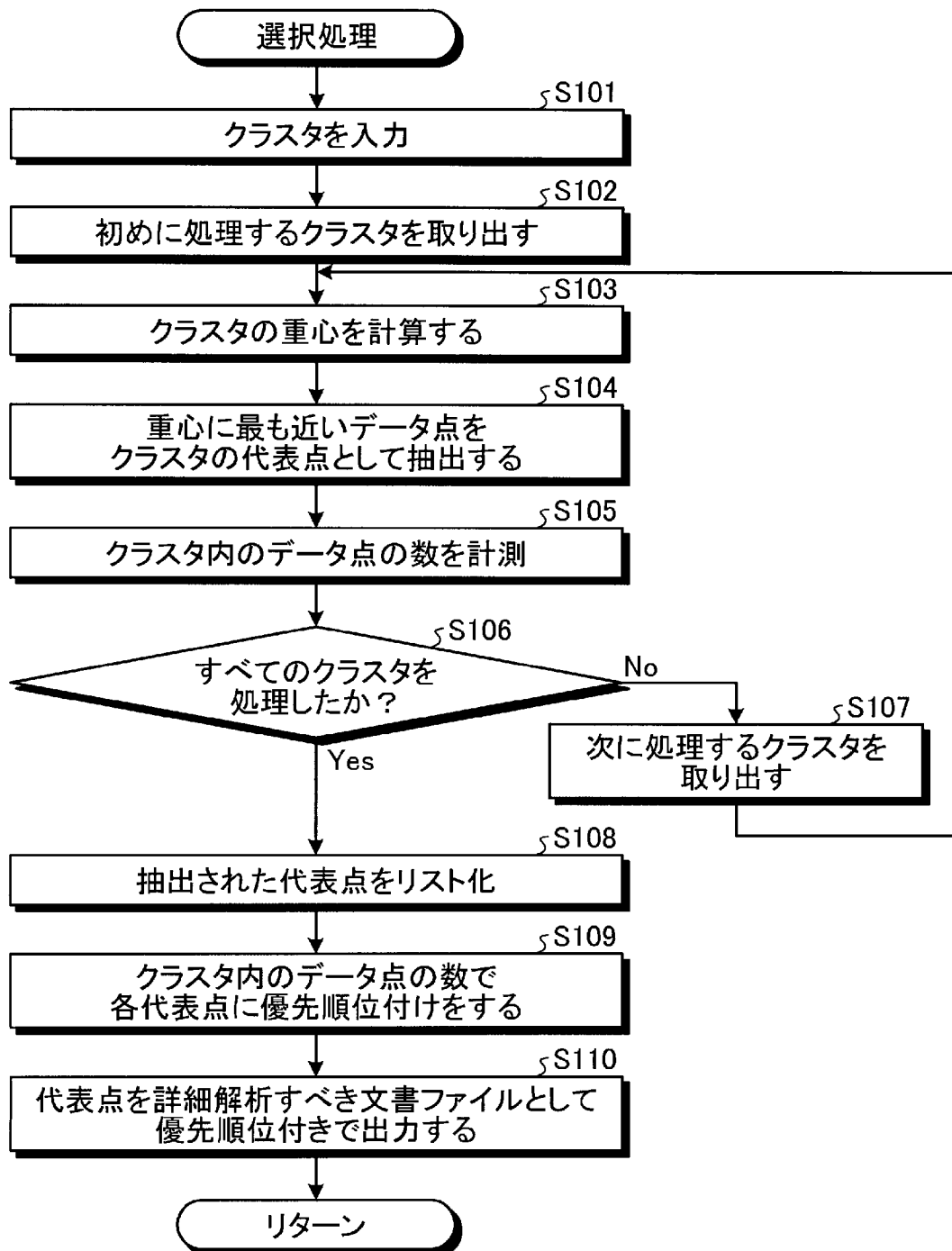
[図10]



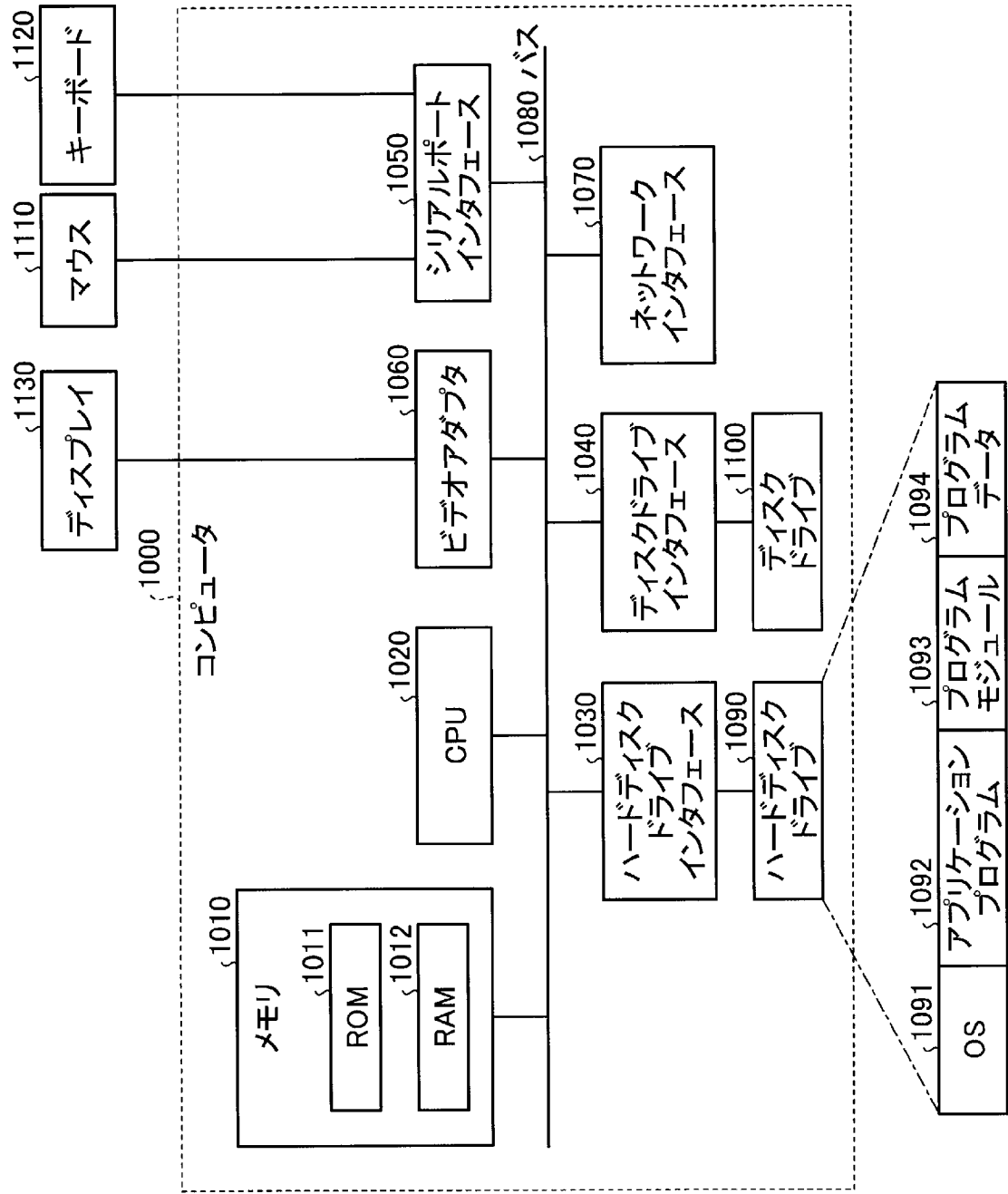
[図11]



[図12]



[図13]



# INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/037407

## A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G06F17/30 (2006.01) i, G06F21/56 (2013.01) 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. G06F17/20-17/30, G06F21/56

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 1971-2017

Registered utility model specifications of Japan 1996-2017

Published registered utility model applications of Japan 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. |
|-----------|--------------------------------------------------------------------------------------------------------------|-----------------------|
| A         | JP 2017-4123 A (NIPPON TELEGRAPH AND TELEPHONE) 05 January 2017, entire text, all drawings<br>(Family: none) | 1-8                   |
| A         | JP 2002-351686 A (SONY CORPORATION) 06 December 2002, entire text, all drawings<br>(Family: none)            | 1-8                   |
| A         | JP 9-223134 A (FUJI XEROX CO., LTD.) 26 August 1997, entire text, all drawings<br>(Family: none)             | 1-8                   |

|                                                                                     |                                                   |
|-------------------------------------------------------------------------------------|---------------------------------------------------|
| <input type="checkbox"/> Further documents are listed in the continuation of Box C. | <input type="checkbox"/> See patent family annex. |
|-------------------------------------------------------------------------------------|---------------------------------------------------|

|                                                                                                                                                                         |                                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| * Special categories of cited documents:                                                                                                                                | "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                                              |
| "A" document defining the general state of the art which is not considered to be of particular relevance                                                                | "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                                                                     |
| "E" earlier application or patent but published on or after the international filing date                                                                               | "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 |
| "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) | "&" document member of the same patent family                                                                                                                                                                                                    |
| "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                                                                  |                                                                                                                                                                                                                                                  |

|                                                                                                                          |                                                                  |
|--------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|
| Date of the actual completion of the international search<br>25.12.2017                                                  | Date of mailing of the international search report<br>09.01.2018 |
| Name and mailing address of the ISA/<br>Japan Patent Office<br>3-4-3, Kasumigaseki, Chiyoda-ku,<br>Tokyo 100-8915, Japan | Authorized officer<br><br>Telephone No.                          |

## 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, G06F21/56(2013.01)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. G06F17/20-17/30, G06F21/56

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<br>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<br>請求項の番号 |
|-----------------|-------------------------------------------------------|----------------|
| A               | JP 2017-4123 A（日本電信電話株式会社）2017.01.05, 全文, 全図（ファミリーなし） | 1-8            |
| A               | JP 2002-351686 A（ソニー株式会社）2002.12.06, 全文, 全図（ファミリーなし）  | 1-8            |
| A               | JP 9-223134 A（富士ゼロックス株式会社）1997.08.26, 全文, 全図（ファミリーなし） | 1-8            |

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

25.12.2017

国際調査報告の発送日

09.01.2018

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）

吉田 誠

5M

3659

電話番号 03-3581-1101 内線 3599