

(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
2017年1月19日(19.01.2017)



(10) 国際公開番号

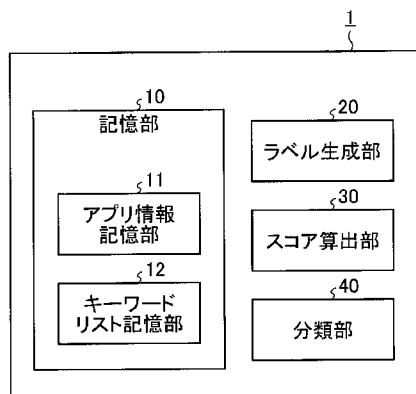
WO 2017/010350 A1

- (51) 国際特許分類:
G06F 21/62 (2013.01) G06F 21/56 (2013.01)
- (21) 国際出願番号: PCT/JP2016/069881
- (22) 国際出願日: 2016年7月5日(05.07.2016)
- (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 2015-139910 2015年7月13日(13.07.2015) 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: 秋山 満昭(AKIYAMA, Mitsuaki); 〒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). 八木 毅(YAGI, 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). 森 達哉(MORI, Tatsuya); 〒1698050 東京都新宿区戸塚町1丁目104番地 学校法人早稲田大学内 Tokyo (JP). 渡邊 卓弥(WATANABE, Takuya); 〒1698050 東京都新宿区戸塚町1丁目104番地 学校法人早稲田大学内 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).
- (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, IR, IS, KE, KG, KN, KP, KR, 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), ユーロッパ (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).

[続葉有]

(54) Title: SOFTWARE ANALYSIS SYSTEM, SOFTWARE ANALYSIS METHOD, AND SOFTWARE ANALYSIS PROGRAM

(54) 発明の名称: ソフトウェア解析システム、ソフトウェア解析方法およびソフトウェア解析プログラム



- 10 Storage unit
11 Application information storing unit
12 Key word list storing unit
20 Label generating unit
30 Score calculating unit
40 Classifying unit

(57) Abstract: In the present invention, on the basis of the code of a plurality of applications, it is determined whether the plurality of applications access, with permission, prescribed privacy information, and a first label is attached to the applications determined to access, and a second label is attached to the applications determined not to access. A score is calculated for each vocabulary item so that the score is higher for vocabulary items a greater number of which are included in the explanatory text of the applications to which the first label is attached than in the explanatory text of the applications to which the second label is attached. A prescribed number of vocabulary items are extracted from vocabulary items having a higher score. The application in which the extracted vocabulary items are included in the explanatory text thereof is classified as the application referring to the permission.

(57) 要約: 複数のアプリケーションのコードに基づき、複数のアプリケーションがプライバシー情報にアクセスするか否かを判定し、アクセスすると判定されたアプリケーションに第1のラベル、アクセスしないと判定されたアプリケーションに第2のラベルを付与する。第2のラベルを付与されたアプリケーションの説明文のテキストよりも、第1のラベルを付与されたアプリケーションの説

明文のテキストに多く含まれる単語のスコアが高くなるよう単語各々のスコアを算出し、スコアが上位から所定数の単語を抽出する。抽出した単語が説明文のテキストに含まれるアプリケーションをパーミッションに言及するアプリケーションとして分類する。



WO 2017/010350 A1

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称：

ソフトウェア解析システム、ソフトウェア解析方法およびソフトウェア解析プログラム

技術分野

[0001] 本発明は、ソフトウェア解析システム、ソフトウェア解析方法およびソフトウェア解析プログラムに関する。

背景技術

[0002] スマートフォン等のモバイル端末の普及に伴い、ユーザがモバイル端末上でダウンロードして利用できるアプリケーションが広く利用されるようになっている。Android（登録商標）やiOS（登録商標）等が搭載されたモバイル端末で動作するアプリケーションは、ユーザのプライバシー情報を処理することが多い。ユーザのプライバシー情報は、モバイル端末に内蔵されたセンサ等のハードウェアが取得するか、または、ユーザ自身がモバイル端末に入力する。プライバシー情報とはたとえば、ユーザの位置情報、モバイル端末の個体識別番号、アドレス帳などユーザ個人に紐付く情報である。

[0003] モバイル端末で動作するアプリケーションはプライバシー情報を処理する際、当該プライバシー情報を適切に取り扱うことが求められる。また、モバイル端末で動作するアプリケーションは、どのようなプライバシー情報にアクセスするか、ユーザに開示することが求められる。

[0004] かかる要求を実現するため、多くのモバイル端末プラットフォームが、パーミッションと呼ばれる機能を導入している。パーミッションとは、プライバシー情報へのアクセスを制御するために導入された機能である。

[0005] たとえば、位置情報を使用するアプリケーションは、当該アプリケーションがモバイル端末にインストールされるときに、ユーザに対して位置情報のパーミッションを使用する許可を求める。ユーザがパーミッションの使用を許可すると、当該アプリケーションがモバイル端末にインストールされる。

ユーザがパーミッションの使用を許可しない場合、当該アプリケーションはモバイル端末にインストールされない。パーミッションを利用することにより、ユーザは、アプリケーションが使用する可能性があるプライバシー情報を予め知ることができる。

先行技術文献

非特許文献

[0006] 非特許文献1: R. Pandita, X. Xiao, W. Yang, W. Enck, and T. Xie, "WHYPER: Towards Automating Risk Assessment of Mobile Applications", in Proc. of the 22nd USENIX Security Symposium, pp. 527-542, 2013

非特許文献2: 渡邊 卓弥、秋山 満昭、酒井 哲也、鷺崎 弘宜、森 達哉、「Androidアプリの説明文とプライバシー情報アクセスの相関分析」、コンピュータセキュリティシンポジウム2014論文集、vol. 2014, No. 2, pp. 590-597, Oct 2014

発明の概要

発明が解決しようとする課題

[0007] しかしながら、従来のパーミッションの機能によって提供される情報は、多くの場合ユーザに注意されない（非特許文献1，2参照）。また、パーミッション機能によって提供される情報は、専門的な知識がなければ理解しにくい。また、パーミッション機能によって提供される情報は、ユーザが直感的に把握しにくい。このため、ユーザが気づかないうちに、アプリケーションがプライバシー情報を取得してしまうことも考えられる。

[0008] かかる状況に鑑み、アプリケーションの説明文中にプライバシー情報にアクセスすることが記載されているか否かを自動判定する技術が提案されている。アプリケーションの説明文はあらゆるマーケットに設定されており、アプリケーションを検索する際の検索対象ともなる。たとえば、非特許文献1は、モバイル端末のアプリケーションの説明文のテキスト（以下、説明文テ

キストまたはテキストとも呼ぶ。)を解析して、テキスト中のどの部分がプライバシー情報へのアクセスに相当するか、特定する技術を提案している。また、非特許文献2は、テキスト解析とアプリケーションのコード解析とを組み合わせ、コード解析から示唆されるプライバシー情報へのアクセスと、テキスト中の記述とに矛盾がある場合に矛盾があるアプリケーションを抽出する技術を提案している。

[0009] しかし、非特許文献1, 2に記載の技術における説明文の解析では、人間が予めアプリケーションの説明文テキストを読み、特定のプライバシー情報にアクセスする旨の記載があるかないかを解読する。そして、解読結果に基づき、アプリケーションにラベルを付与する。しかし、プライバシー情報は多岐にわたり、機械学習等の統計的方法でテキストの分類を試みる場合、大量のテキストを準備してラベルを付与することになる。一般的に、手動でデータにラベルを付与するという作業はコストがかかる。また、大規模なデータに対して手動でラベルを付与するという手法はスケーラビリティに欠ける。

[0010] 開示の技術は、上記に鑑みてなされたものであって、アプリケーションのコードおよび説明文に基づき、アプリケーションによるユーザのプライバシー情報へのアクセスの有無がユーザに対して開示されているか否かを自動的に解析することを可能にするソフトウェア解析システム、ソフトウェア解析方法およびソフトウェア解析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

[0011] 開示するソフトウェア解析システム、ソフトウェア解析方法およびソフトウェア解析プログラムは、複数のアプリケーションのコードに基づき、当該複数のアプリケーションがパーミッションにより所定のプライバシー情報にアクセスするか否かを判定する。そして、開示するソフトウェア解析システム、ソフトウェア解析方法およびソフトウェア解析プログラムは、所定のプライバシー情報にアクセスすると判定されたアプリケーションに第1のラベルを付与し、所定のプライバシー情報にアクセスしないと判定されたアプリ

ケーションに第1のラベルと異なる第2のラベルを付与する。開示するソフトウェア解析システム、ソフトウェア解析方法およびソフトウェア解析プログラムは、第2のラベルを付与されたアプリケーションの説明文のテキストよりも、第1のラベルを付与されたアプリケーションの説明文のテキストに多く含まれる単語のスコアが高くなるよう複数のアプリケーションの説明文に含まれる単語各々のスコアを算出し、パーミッションに対応付けて、算出されたスコアが上位から所定数の単語を、複数のアプリケーションの説明文から抽出する。さらに、開示するソフトウェア解析システム、ソフトウェア解析方法およびソフトウェア解析プログラムは、抽出した単語が説明文のテキストに含まれるアプリケーションをパーミッションに言及するアプリケーションとして分類し、抽出した単語が説明文のテキストに含まれないアプリケーションをパーミッションに言及しないアプリケーションとして分類することを特徴とする。

発明の効果

- [0012] 開示するソフトウェア解析システム、ソフトウェア解析方法およびソフトウェア解析プログラムは、アプリケーションのコードおよび説明文に基づき、アプリケーションによるユーザのプライバシー情報へのアクセスの有無がユーザに対して開示されているか否かを自動的に解析することを可能にするという効果を奏する。

図面の簡単な説明

- [0013] [図1]図1は、実施形態に係るソフトウェア解析システムの構成の一例を示す概略図である。
- [図2]図2は、アプリ情報記憶部に格納される情報の構成の一例を説明するための図である。
- [図3]図3は、キーワードリスト記憶部に格納される情報の構成の一例を説明するための図である。
- [図4]図4は、実施形態に係るソフトウェア解析システムが実行する処理の流れの一例を示すフローチャートである。

[図5]図 5 は、実施形態に係るラベル生成処理の流れの一例を示すフローチャートである。

[図6]図 6 は、実施形態に係るスコア算出処理（キーワード抽出処理）の流れの一例を示すフローチャートである。

[図7]図 7 は、実施形態に係る分類処理の流れの一例を示すフローチャートである。

[図8]図 8 は、実施形態に係るソフトウェア解析システムが実行する処理の流れの一例を説明するためのシーケンス図である。

[図9]図 9 は、変形例に係るソフトウェア解析システムの構成の一例を示す概略図である。

[図10]図 10 は、変形例に係るアプリ情報記憶部に格納される情報の構成の一例を示す図である。

[図11]図 11 は、変形例に係るソフトウェア解析システムが実行する処理の流れの一例を示すフローチャートである。

[図12]図 12 は、実施形態に係るソフトウェア解析プログラムを実行するコンピュータを示す図である。

発明を実施するための形態

[0014] 以下に、開示するシステム、方法およびプログラムの実施形態を図面に基づいて詳細に説明する。なお、この実施形態によりこの発明が限定されるものではない。また、各実施形態は適宜組み合わせることができる。

[0015] [ソフトウェア解析システムの構成の一例]

図 1 は、実施形態に係るソフトウェア解析システム 1 の構成の一例を示す概略図である。図 1 に示すソフトウェア解析システム 1 は、たとえば、ユーザがモバイル端末にアプリケーションをダウンロードするためにアクセスするダウンロードサービスの運営者が運用する、アプリケーションの事前審査を行うサーバである。なお、以下の記載中、「アプリケーション」は「アプリ」とも呼ぶ。

[0016] ソフトウェア解析システム 1 は、記憶部 10 と、ラベル生成部 20 と、ス

コア算出部 30 と、分類部 40 と、を備える。

[0017] 記憶部 10 はたとえば、半導体メモリ素子や記憶装置である。例えば、半導体メモリ素子としては、VRAM (Video Random Access Memory)、RAM (Random Access Memory)、ROM (Read Only Memory) やフラッシュメモリ (flash memory) などを用いることができる。また、記憶装置としては、ハードディスク、光ディスクなどの記憶装置を用いることができる。

[0018] 記憶部 10 は、ソフトウェア解析システム 1 による解析処理の対象となるアプリケーションに関する情報を記憶する。また、記憶部 10 は、ソフトウェア解析システム 1 による解析処理の結果得られる情報を記憶する。

[0019] 記憶部 10 は、アプリ情報記憶部 11 と、キーワードリスト記憶部 12 と、を備える。図 2 は、アプリ情報記憶部 11 に格納される情報の構成の一例を説明するための図である。また、図 3 は、キーワードリスト記憶部 12 に格納される情報の構成の一例を説明するための図である。

[0020] 図 2 に示すように、アプリ情報記憶部 11 は、「アプリ ID」に対応付けて、「コード」と、「説明文テキスト」と、「プライバシー情報」と、「ラベル」と、「分類結果」と、を記憶する。

[0021] 「アプリ ID」はソフトウェア解析システム 1 の解析処理の対象となるアプリケーションを一意に識別するための識別子である。「コード」は、当該アプリケーションのコードである。たとえば、アプリケーションのソースコードである。「説明文テキスト」は、当該アプリケーションの説明文のテキストである。「プライバシー情報」は、パーミッションによりアプリケーションのアクセスが許可されるユーザのプライバシー情報の種類を示す。プライバシー情報としては、たとえば、GPS (Global Positioning System) 情報や、カレンダー、カメラ機能、アドレス帳等がある。

[0022] 「ラベル」は、ソフトウェア解析システム 1 によるコード解析の結果にもとづき、ソフトウェア解析システム 1 が各アプリケーションに付与するラベルである。また、「ラベル」は、アプリケーションが所定のプライバシー情

報にアクセスするか否かの判定結果を示す暫定的なラベルである。ラベルには、「言及あり」と「言及なし」の2種類がある。「言及あり」は、アプリケーションが所定のプライバシー情報にアクセスすることを示す情報がコード中に存在することを示す。また、「言及なし」は、アプリケーションが所定のプライバシー情報にアクセスすることを示す情報がコード中に存在しないことを示す。「分類結果」は、ソフトウェア解析システム1が、説明文テキストを解析して抽出したキーワードが、各アプリケーションの説明文テキストに含まれるか否かにもとづき、アプリケーションを分類した結果を示す。「分類結果」には、「言及あり」と「言及なし」の2種類がある。「言及あり」は、アプリケーションの説明文テキストにキーワードが含まれることを示す。「言及なし」は、アプリケーションの説明文テキストにキーワードが含まれないことを示す。「ラベル」および「分類結果」は、プライバシー情報各々に対応づけて格納される。なお、ラベルの生成処理および分類結果を得るための分類処理については後述する。なお、図2では、プライバシー情報とパーミッションとが一意に対応づけられることを前提として、ラベルおよび分類結果とプライバシー情報とを対応付けて記憶するが、プライバシー情報の代わりにパーミッションに対応づけて記憶するように構成してもよい。

- [0023] 図2中、たとえば、アプリID「A P O O 1」のアプリケーションに対応づけて、当該アプリケーションの説明文テキスト「〜〜〜」と、対応するコード「〜〜〜」と、が記憶されている。また、当該アプリケーションに対応づけて、当該アプリケーションが3種類のプライバシー情報各々にアクセスするか否かを示すラベルと分類結果とが記憶されている。「プライバシー情報、カレンダー」については、「ラベル、言及あり」が記憶されている。これは、コード解析によれば、アプリID「A P O O 1」のアプリケーションはユーザのカレンダーにアクセスする、ということを示す。また図2中、アプリID「A P O O 1」に対応付けて「プライバシー情報、アドレス帳」、「分類結果、言及なし」が記憶されている。これは、アプリID「A P O O

1」で特定されるアプリケーションの説明文テキストには、プライバシー情報「アドレス帳」に対応するキーワードが含まれないことを示す。

[0024] 次に、図3を参照して、キーワードリスト記憶部12に記憶される情報について説明する。図3に示すように、キーワードリスト記憶部12は、「プライバシー情報」と、「機能（パーミッション）」と、「キーワード」と、を記憶する。「プライバシー情報」は、アプリ情報記憶部11に記憶される「プライバシー情報」と同様である。「機能（パーミッション）」は、対応するプライバシー情報へのアクセスを許可するパーミッションの名称である。パーミッションとはたとえばAndroid（登録商標）について規定されている「READ_CONTACTS」、「READ_CALENDER」などである。「キーワード」は、ソフトウェア解析システム1によりアプリケーションの説明文テキストから抽出されるキーワードである。キーワードを抽出する処理の詳細については後述する。

[0025] 図3の例では、「プライバシー情報、GPS情報」に対応づけて、「機能（パーミッション）、Location」と、「キーワード、gps, location, map」とが記憶されている。これは、パーミッション「Location」の機能によってアクセスが許可されるプライバシー情報は、GPS情報であることを示す。また、パーミッション「Location」に対応するキーワードとして「gps, location, map」の3つの単語が抽出されたことを示す。なお、図3の例では、各プライバシー情報および機能（パーミッション）に対応づけて3つのキーワードが記憶されている。ただし、各プライバシー情報および機能（パーミッション）に対応づけられるキーワードの数は3つに限定されない。

[0026] 図1に戻り、さらにソフトウェア解析システム1の構成の一例につき説明する。ラベル生成部20は、記憶部10に記憶されるアプリケーションの情報にもとづき、コード解析の結果、各アプリケーションがユーザのプライバシー情報にアクセスすると判定されたか否かを示すラベルを生成する。

[0027] スコア算出部30は、ラベル生成部20により生成されたラベルを用いて、アプリケーションの説明文テキストを解析し、説明文テキストに含まれる

単語各々のスコアを算出する。スコア算出部 30 は、パーミッションごとに各単語のスコアを算出する。スコア算出部 30 は、ラベル「言及なし」のアプリケーションの説明文のテキストよりも、ラベル「言及あり」のアプリケーションの説明文のテキストに多く含まれる単語のスコアが高くなるように、説明文テキストに含まれる単語各々のスコアを算出する。ただし、算出手法は特に限定されず、たとえば、オッズ比や適合性フィードバックを用いる手法を適用することができる。これらの手法を用いて、たとえば、ラベル「言及あり」のアプリケーションの総数、説明文のテキストが所定の単語を含むアプリケーションの総数、ラベル「言及あり」のアプリケーションのうち説明文のテキストが所定の単語を含むアプリケーションの総数等に基づきスコアを算出することができる。

[0028] また、スコア算出部 30 が算出するスコアの算出手法は、ラベル「言及あり」のアプリケーションの説明文テキストに頻出する単語であって、ラベル「言及なし」のアプリケーションの説明文テキストにはほとんど出現しない単語ほど高い値となるように設定する。つまり、ラベル「言及あり」のアプリケーションの説明文テキストにおける出現頻度が、ラベル「言及なし」のアプリケーションの説明文テキストにおける出現頻度よりも高い単語ほどスコアが高くなるよう、スコアの算出手法を設定する。また、スコアの算出手法は、API (Application Programming Interface) を使用するアプリケーションの説明文テキストに頻出する単語であって、API を使用していないアプリケーションの説明文テキストにはほとんど出現しない単語ほど高い値となるように設定する。

[0029] スコア算出部 30 は、パーミッションに対応付けてスコアが高い順に上位所定数の単語を、パーミッションに対応するキーワードとして抽出する。キーワードとは、対応するプライバシー情報にアクセスするアプリケーションの説明文テキストに出現すると予想される単語である。スコア算出部 30 が抽出したキーワードは、プライバシー情報およびパーミッションに対応づけて、キーワードリスト記憶部 12 に記憶される。

[0030] 分類部40は、各アプリケーションの説明文テキストが、抽出されたキーワードを含むか否かを判定する。そして、分類部40は、キーワードを含む説明文テキストのアプリケーションを、「言及あり」のアプリケーションとして分類する。また、分類部40は、キーワードを含まない説明文テキストのアプリケーションを、「言及なし」のアプリケーションとして分類する。分類部40による分類結果は、各アプリケーションおよびプライバシー情報に対応づけて、アプリ情報記憶部11に記憶される。

[0031] [ソフトウェア解析システムによる解析処理の流れの一例]

図4は、実施形態に係るソフトウェア解析システム1が実行する処理の流れの一例を示すフローチャートである。図4の処理の前提として、ソフトウェア解析システム1は予め、複数のアプリケーションに関する情報を取得し、記憶部10のアプリ情報記憶部11に記憶しているものとする。図4に示すように、ソフトウェア解析システム1による解析処理は、ラベル生成部20によるラベル生成（コード解析）（ステップS21）、スコア算出部30によるスコア算出（キーワード抽出）（ステップS22）、分類部40による分類（ステップS23）の各処理を含む。分類部40による分類が完了すると処理は終了する。以下、各部による処理の流れについて説明する。

[0032] [ラベル生成処理の流れの一例]

ラベル生成部20は、アプリ情報記憶部11に記憶されるアプリケーションの情報にもとづき、各アプリケーションのラベルを生成する。図5は、実施形態に係るラベル生成処理の流れの一例を示すフローチャートである。図5に示すように、ラベル生成部20はまず、解析の対象として、1つのアプリケーションと、1つのプライバシー情報を選択する（ステップS51）。たとえば、ラベル生成部20は、アプリID「AP001」のアプリケーションを選択する。また、ラベル生成部20は、プライバシー情報としてカメラ機能を選択する。ラベル生成部20によるアプリケーションおよびプライバシー情報の選択の順序については、あらかじめ決定してソフトウェア解析システム1に記憶させておいてもよいし、選択するアプリケーションおよび

プライバシー情報をユーザが指定するようにしてもよい。

- [0033] そして、ラベル生成部20は、選択したアプリケーションのコードをアプリ情報記憶部11から読み出す。アプリケーションのコードは、ソースコードが取得できる場合はソースコードを取得してアプリ情報記憶部11に記憶しておく。また、ソースコードが取得できない場合は、リバースエンジニアリング等の手法を用いて逆アセンブリしたコードをアプリ情報記憶部11に記憶しておく。
- [0034] 次に、ラベル生成部20は、読み出したコード中に、パーミッションの宣言が含まれているか否かを判定する（ステップS52）。ラベル生成部20はたとえば、Android（登録商標）を用いたアプリケーションの場合であれば、AndroidManifest.xmlの中でパーミッションが宣言されているか否かを判定する。
- [0035] ラベル生成部20は、パーミッションが宣言されていると判定した場合（ステップS52、Yes）、次に、選択したプライバシー情報（すなわちパーミッション）に対応するAPIがコード中に含まれているか否かを判定する（ステップS53）。さらに、APIが含まれていると判定した場合（ステップS53、Yes）、ラベル生成部20は、コールグラフを解析して、当該APIが実際に呼び出し可能であるか否かを判定する（ステップS54）。そして、コールグラフを解析した結果、呼び出し可能と判定した場合（ステップS54、Yes）、ラベル生成部20は、当該アプリケーションにラベル「言及あり」を生成する（ステップS55）。
- [0036] 他方、パーミッションが宣言されていないと判定した場合（ステップS52、No）には、ラベル生成部20は、当該アプリケーションに「言及なし」のラベルを付与する（ステップS56）。APIは含まれていないと判定した場合（ステップS53、No）およびコールグラフ解析から呼び出しできないと判定した場合（ステップS54、No）も同様に、ラベル生成部20は、「言及なし」のラベルを生成する。これによってラベル生成処理が終了する。生成されたラベルは、アプリケーションおよびプライバシー情報に

対応づけて、アプリ情報記憶部 11 に記憶される。

[0037] ラベル生成部 20 は、図 5 に示す処理を各アプリケーションおよびプライバシー情報について繰り返し実行することにより、複数のアプリケーションおよびプライバシー情報に対応するラベルを生成し、アプリ情報記憶部 11 に記憶する。

[0038] なお、図 5 のように 3 段階の判定を経てラベルを付与するのは、コード中にパーミッションの宣言や対応する API がある場合でも、実際はプライバシー情報にアクセスしないコードとなっていることがあるからである。

[0039] [スコア算出処理（キーワード抽出処理）の流れの一例]

スコア算出部 30 は、ラベルが付与されたアプリケーションおよびプライバシー情報を用いて、プライバシー情報にアクセスするアプリケーションを、説明文テキストに基づいて特定するためのキーワードを抽出する。

[0040] 図 6 は、実施形態に係るスコア算出処理（キーワード抽出処理）の流れの一例を示すフローチャートである。まず、スコア算出部 30 は、ラベル生成処理が終了した複数のアプリケーションに対応する説明文テキストを、アプリ情報記憶部 11 から読み出す（ステップ S 61）。そして、スコア算出部 30 は、パーミッションを選択する（ステップ S 62）。さらに、スコア算出部 30 は、読み出した説明文テキストを解析して、説明文テキストに含まれる単語を一つ選択する（ステップ S 63）。

[0041] そして、スコア算出部 30 は、選択した単語のスコアを算出する（ステップ S 64）。以下に、適合フィードバックを用いてスコアを算出する例を説明する。一例として、スコア算出部 30 は、以下の式（1）を用いてスコアを算出する。式（1）は、総数 N 個（ N は自然数）のアプリケーションを解析対象とし、そのうち、 i 番目（ i は、1 以上 N 以下の自然数）のアプリケーション A_i に含まれる単語 w_i のスコアを算出するための式である。

[数1]

$$RW(w_i) = \log \frac{(r_i + 0.5)(N - n_i - R + r_i + 0.5)}{(n_i - r_i + 0.5)(R - r_i + 0.5)} \quad \dots (1)$$

ただし、式（１）中、 R は、ラベル「言及あり」を付与されたアプリケーションの総数であり、 r_i は単語 w_i を説明文テキストに含みラベル「言及あり」を付与されたアプリケーションの総数である。また、 N は解析対象とするアプリケーションの総数であり、 n_i は単語 w_i を説明文テキストに含むアプリケーションの総数である。また、式（１）においては、分子および分母がゼロになることを避けるため、 0.5 を加算している。ただし、加算する値は 0.5 に限定されず、たとえば 0.1 でも 0.6 でもよい。

[0042] スコア算出部３０は次に、すべての単語のスコアを算出したか否かを判定する（ステップＳ６５）。すべての単語のスコアを算出していないと判定した場合（ステップＳ６５、 No ）、スコア算出部３０は次に処理対象とする単語を選択する（ステップＳ６６）。そして、処理はステップＳ６４に戻る。他方、すべての単語のスコアを算出したと判定した場合（ステップＳ６５、 Yes ）、スコア算出部３０は、パーミッションごとに算出したスコアの降順に単語をソートし、上位所定数の単語を抽出する（ステップＳ６７）。たとえば、スコア算出部３０は、上位３個の単語を抽出する。ここで、抽出する単語の数を３とするのは、本発明者らの実験結果から、抽出する単語数を３とした場合に、最終的に良好な分類精度が得られているためである。ただし、抽出する単語数は他の数に設定してもよい。これによってスコア算出処理（キーワード抽出処理）が終了する。

[0043] なお、上の例では、 $relevance\ weight$ を用いてスコアを算出する手法を採用した。しかし、スコアを算出するための式は上記の例に限定されない。スコアを算出するための式は、ある集合のアプリケーションに含まれやすく、他の集合のアプリケーションに含まれにくい、単語の性質を数値化することができればよい。

[0044] たとえば、オッズ比をスコアとして採用することができる。まず、所定の単語がラベル「言及あり」のアプリケーションの集合に含まれる確率を p 、ラベル「言及なし」のアプリケーションの集合に含まれる確率を q とする。そして、 $(p / (1 - p))$ を $(q / (1 - q))$ で除してオッズ比を算出

する。算出されるオッズ比が大きいほど、所定の単語がラベル「言及あり」のアプリケーションの集合に含まれやすく、ラベル「言及なし」のアプリケーションの集合に含まれにくいことを意味する。

[0045] たとえば、単語「SNS」がラベル「言及あり」のアプリケーションの集合に含まれる確率 p が0.9であり、ラベル「言及なし」のアプリケーションの集合に含まれる確率 q が0.1であれば、上の式に基づいて算出されるオッズ比は、8.1となる。同様にして他の単語についてもオッズ比を算出しオッズ比をスコアとする。そして、スコアが高い方から順に単語を順位づける。最後に、順位が上の単語から順に「言及あり」のアプリケーションの集合により含まれやすいキーワードとして抽出すればよい。

[0046] [分類処理の流れの一例]

分類部40は、スコア算出部30が抽出したキーワードと、アプリケーションの説明文テキストに基づき、各アプリケーションがプライバシー情報にアクセスするか否かに基づきアプリケーションを分類する。図7は、実施形態に係る分類処理の流れの一例を示すフローチャートである。

[0047] 分類部40はまず、分類対象とするアプリケーションの説明文テキストを、アプリ情報記憶部11から読み出す。そして、分類部40は、当該説明文テキストに含まれるすべての単語を抽出する（ステップS71）。そして、分類部40は、スコア算出部30が抽出したキーワードと一致する単語が、説明文テキストから抽出された単語に含まれるか否かを判定する（ステップS72）。一致する単語が含まれると判定した場合（ステップS72、Yes）、分類部40は、当該アプリケーションを、キーワードに対応するパーミッションに「言及あり」のアプリケーションとして分類する（ステップS73）。ここで、分類「言及あり」は、キーワードに対応するパーミッションによりアクセスが許可されるプライバシー情報にアクセスするアプリケーションであると判定されたことを示す。他方、一致する単語が含まれないと判定した場合（ステップS72、No）、分類部40は、当該アプリケーションをキーワードに対応するパーミッションに「言及なし」のアプリケーション

ョンとして分類する（ステップS 7 4）。ここで、分類「言及なし」は、キーワードに対応するパーミッションによりアクセスが許可されるプライバシー情報にアクセスしないアプリケーションであると判定されたことを示す。これで、分類処理は終了する。

[0048] 図8は、実施形態に係るソフトウェア解析システム1が実行する処理の流れの一例を説明するためのシーケンス図である。図8に示す処理は、図4乃至図7を参照して説明した処理の流れに対応する。

[0049] 図8に示すように、ソフトウェア解析システム1においては、まず、解析の対象とするアプリケーションの情報や説明文テキストが用意される（図8の（1））。たとえば、ソフトウェア解析システム1は、ネットワークを介して、アプリケーションのデータを取得し、記憶部10に記憶する。

[0050] そして、ラベル生成部20は、記憶部10に記憶されたアプリケーションのコードを参照して（図8の（2））、当該コードの解析を実行する（図8の（3））。ラベル生成部20は、コード解析の結果にもとづきラベルを生成し、記憶部10に記憶する（図8の（4））。

[0051] スコア算出部30は、記憶部10に記憶されたアプリケーションの情報とともにラベルを取得し（図8の（5））、スコア算出処理（キーワード生成処理）を実行する（図8の（6））。スコア算出部30の処理の結果、キーワードリストが生成され、記憶部10に記憶される。次に、分類部40は、記憶部10からキーワードリストを取得する（図8の（7））。さらに、分類部40は、記憶部10から、説明文テキストを取得する（図8の（8））。分類部40は、取得したキーワードリストと説明文テキストとに基づき、テキスト分類を実行する（図8の（9））。そして、分類部40の処理の結果得られる分類結果を、記憶部10に記憶する（図8の（10））。

[0052] [実施形態の効果]

上記のように、上記実施形態に係るソフトウェア解析システムは、ラベル生成部と、スコア算出部と、分類部と、を備える。そして、ラベル生成部は、複数のアプリケーションのコードに基づき、当該複数のアプリケーション

がパーミッションにより所定のプライバシー情報にアクセスするか否かを判定する。また、ラベル生成部は、所定のプライバシー情報にアクセスすると判定されたアプリケーションに第1のラベル（「言及あり」）を付与し、所定のプライバシー情報にアクセスしないと判定されたアプリケーションに第1のラベルと異なる第2のラベル（「言及なし」）を付与する。スコア算出部は、第2のラベルを付与されたアプリケーションの説明文のテキストよりも、第1のラベルを付与されたアプリケーションの説明文のテキストに多く含まれる単語のスコアが高くなるよう複数のアプリケーションの説明文に含まれる単語各々のスコアを算出し、パーミッションに対応付けて、算出されたスコアが上位から所定数の単語を、複数のアプリケーションの説明文から抽出する。分類部は、スコア算出部が抽出した単語が説明文のテキストに含まれるアプリケーションをパーミッションに言及するアプリケーションとして分類し、スコア算出部が抽出した単語が説明文のテキストに含まれないアプリケーションをパーミッションに言及しないアプリケーションとして分類する。

[0053] このため、実施形態のソフトウェア解析システムは、コード解析と説明文のテキスト解析とを組み合わせ、アプリケーションによるプライバシー情報へのアクセスの有無と、アプリケーションの説明文におけるプライバシー情報へのアクセスの言及の有無を精度高く判定することができる。また、実施形態のソフトウェア解析システムは、説明文のテキスト解析を自動化することができ、テキスト解析の処理効率およびコストを向上させることができる。また、実施形態のソフトウェア解析システムは、プライバシー情報にアクセスするアプリケーションの説明文に特徴的に含まれる単語の抽出を自動的に実行することができ、処理効率およびコストを向上させることができる。このように、実施形態のソフトウェア解析システムによれば、アプリケーションのコードおよび説明文に基づき、アプリケーションによるユーザのプライバシー情報へのアクセスの有無がユーザに対して開示されているか否かを自動的に解析することができる。

[0054] [変形例]

上記実施形態では、ソフトウェア解析システム 1 は、各アプリケーションのコードと説明文テキストとに基づき、当該アプリケーションが所定のプライバシー情報にアクセスするか否かを判定した。また、上記実施形態のソフトウェア解析システム 1 は、説明文テキストに基づき、各アプリケーションがプライバシー情報へのアクセスを説明文において開示しているか否かを分類した。以下で、変形例として、ソフトウェア解析システムが、新しく検知されたアプリケーションに対して解析をおこなって分類結果を取得し、分類結果に基づき、悪性アプリケーションと良性アプリケーションとを特定する例を説明する。変形例のソフトウェア解析システムはまた、新しく検知したアプリケーションが悪性アプリケーションである場合にユーザに警告するよう構成される。

[0055] 図 9 は、変形例に係るソフトウェア解析システム 1 A の構成の一例を示す概略図である。図 9 に示すように、変形例に係るソフトウェア解析システム 1 A は、記憶部 10 A と、ラベル生成部 20 と、スコア算出部 30 と、分類部 40 と、特定部 50 と、送信部 60 と、を備える。また、記憶部 10 A は、アプリ情報記憶部 11 A と、キーワードリスト記憶部 12 と、を備える。ラベル生成部 20、スコア算出部 30、分類部 40 は、上記実施形態のソフトウェア解析システム 1 が備えるラベル生成部 20、スコア算出部 30、分類部 40 と同様の構成および機能である。また、キーワードリスト記憶部 12 に記憶される情報も上記実施形態のソフトウェア解析システム 1 の場合と同様である。変形例のソフトウェア解析システム 1 A は、特定部 50、送信部 60 を備える点と、アプリ情報記憶部 11 A に記憶される情報とが、上記実施形態のソフトウェア解析システム 1 と異なる。以下の説明中、上記実施形態のソフトウェア解析システム 1 と同様の構成および機能については説明を省略し、相違する点について説明する。

[0056] 特定部 50 は、ラベル生成部 20 が生成したラベルと、分類部 40 の分類結果とを比較し、矛盾するアプリケーションを特定する。たとえば、図 2 中

、アプリID「A P O O 1」のアプリケーションについて記憶される情報中、「プライバシー情報、GPS情報」に対応するラベルは「言及あり」、分類結果は「言及なし」となっている。これは、アプリID「A P O O 1」で特定されるアプリケーションのコード解析の結果によれば、当該アプリケーションはユーザのGPS情報にアクセスすることを示す。また、当該アプリケーションの説明文テキストの解析の結果によれば、当該アプリケーションがユーザのGPS情報にアクセスすることについて説明文では言及されていないことを示す。すなわち、アプリID「A P O O 1」のアプリケーションは、ユーザのプライバシー情報である「GPS情報」にアクセスするアプリケーションであるにも関わらず、ユーザは説明文を読んだだけではそのことが分からない状態で配布されるアプリケーションである。特定部50は、このように、「ラベル」と「分類結果」とが一致しないアプリケーションを特定する。

[0057] 特定部50が特定したアプリケーションについては、アプリ情報記憶部11Aに悪性であることを示すフラグを記憶させる。図10は、変形例に係るアプリ情報記憶部11Aに格納される情報の構成の一例を示す図である。図10の例では、「アプリID、A P O O 1」、「プライバシー情報、GPS情報」、「ラベル、言及あり」、「分類結果、言及なし」に対応づけて、「フラグ、ON」が記憶されている。これは、アプリID「A P O O 1」のアプリケーションはGPS情報にアクセスするにもかかわらず、その旨を説明文に記載しておらず、悪性アプリケーションと推定されることを意味する。

[0058] 送信部60は、特定部50により悪性アプリケーションと特定されたアプリケーションをユーザがダウンロードしようとする際に、当該アプリケーションが悪性アプリケーションと推定される旨の警告を当該ユーザのモバイル端末に送信する。送信部60はたとえば、悪性アプリケーションと推定される旨と、当該アプリケーションによりアクセスされるプライバシー情報と、の通知を送信する。

[0059] 図11は、変形例に係るソフトウェア解析システム1Aが実行する処理の

流れの一例を示すフローチャートである。まず、特定部50は、対応するラベルと、分類結果とを、アプリ情報記憶部11Aから読み出す（ステップS111）。そして、特定部50は、読み出したラベルと分類結果との間に矛盾があるか否かを判定する（ステップS112）。たとえば、ラベルが「言及あり」、分類結果が「言及なし」の場合、特定部50は、矛盾ありと判定する。また、ラベルが「言及なし」、分類結果が「言及なし」の場合、特定部50は、矛盾なしと判定する。

[0060] 矛盾ありと判定した場合（ステップS112、Yes）、特定部50は、当該アプリケーションを悪性アプリと判定する（ステップS113）。そして、特定部50は、アプリ情報記憶部11Aに記憶される当該アプリケーションの情報に対応づけて、悪性アプリケーションである旨を示すフラグをONにする。他方、矛盾なしと判定した場合（ステップS112、No）、特定部50は処理を終了する。

[0061] そして、送信部60は、特定部50が悪性アプリケーションと特定したアプリケーションをダウンロードしたユーザに対して警告を送信する（ステップS114）。これによって、変形例における特定部50および送信部60の処理が終了する。

[0062] [変形例の効果]

このように、変形例に係るソフトウェア解析システム1Aは、ラベル生成部により第1のラベル（「言及あり」）を付与されたアプリケーションであって、分類部によりパーミッションに言及しないアプリケーションに分類されたアプリケーションを、悪性アプリケーションとして特定する特定部を、さらに備える。このため、ソフトウェア解析システム1Aは、コード解析の結果とテキスト解析の結果とに矛盾があるアプリケーションを容易に特定して、悪性アプリケーションとして抽出することができる。

[0063] また、変形例に係るソフトウェア解析システム1Aは、ラベル生成部により第1のラベル（「言及あり」）を付与されたアプリケーションであって、分類部によりパーミッションに言及しないアプリケーションに分類されたア

アプリケーションが、モバイル端末にインストールされる際に、警告情報を送信する送信部をさらに備える。このため、モバイル端末を利用してアプリケーションをダウンロードするユーザに、悪性アプリケーションの可能性があることを警告することができる。

[0064] (その他の実施形態)

これまで本発明の実施形態について説明したが、本発明は上述した実施形態以外にも、その他の実施形態にて実施されてもよい。以下に、その他の実施形態を説明する。

[0065] 上記実施形態では、予めソフトウェア解析システム 1 に 1 以上のアプリケーションの情報が記憶されているものとした。しかし、これに限定されず、ソフトウェア解析システムは、ネットワークを介してアプリケーションを提出するアプリケーション作成者の情報端末等と接続され、当該情報端末から送信されるアプリケーションを受信し解析するように構成してもよい。この場合、特定部 50 による処理の結果を、送信部 60 がアプリケーション作成者の情報端末に送信するように構成してもよい。また、この場合、スコア算出部によるスコアの算出やキーワードの抽出は実行しないように構成してもよい。すなわち、分類部は、すでにキーワードリスト記憶部 12 に記憶されているキーワードを用いて、新しいアプリケーションの分類を実行するものとしてもよい。このように構成すれば、新しいアプリケーションが受信されるたびに、キーワードリストを更新する必要がなく、分類部の処理を迅速に実行することができる。

[0066] また、ソフトウェア解析システム 1 は、ネットワークを介してアプリケーションサービスを提供する 1 以上の事業者のサーバと接続されるように構成してもよい。そして、当該事業者のサーバからソフトウェア解析システム 1 に対して、新しいアプリケーションが適宜送付され、解析対象となるように構成してもよい。

[0067] [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、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。例えば、キーワードリスト記憶部 12 をソフトウェア解析システム 1 の外部の記憶装置に記憶するように構成して他のサーバがキーワードを利用して悪性アプリケーションを特定できるように構成してもよい。

[0068] さらに、各装置にて行なわれる各処理機能は、その全部または任意の一部が、CPU および当該 CPU にて解析実行されるプログラムにて実現され、あるいは、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。

[0069] 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または一部を手動的に行うこともでき、あるいは、手動的に行われるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。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。

[0070] [プログラム]

また、上記実施の形態において説明したソフトウェア解析システム 1, 1A が実行する処理をコンピュータが実行可能な言語で記述したプログラムを作成することもできる。例えば、実施形態に係るソフトウェア解析システム 1, 1A が実行する処理をコンピュータが実行可能な言語で記述したプログラムを作成することもできる。この場合、コンピュータがプログラムを実行することにより、上記実施形態と同様の効果を得ることができる。さらに、かかるプログラムをコンピュータが読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータに読み込ませて実行することにより上記実施形態と同様の処理を実現してもよい。以下に、ソフトウェア解析システム 1, 1A と同様の機能を実現するプログラムを実行するコンピュータの一例を説明する。

[0071] 図12は、ソフトウェア解析プログラムを実行するコンピュータ1000を示す図である。図12に例示するように、コンピュータ1000は、例えば、メモリ1010と、CPU (Central Processing Unit) 1020と、ハードディスクドライバインタフェース1030と、ディスクドライバインタフェース1040と、シリアルポートインタフェース1050と、ビデオアダプタ1060と、ネットワークインタフェース1070とを有し、これらの各部はバス1080によって接続される。

[0072] メモリ1010は、図12に例示するように、ROM (Read Only Memory) 1011及びRAM (Random Access Memory) 1012を含む。ROM 1011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、図12に例示するように、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、図12に例示するように、ディスクドライブ1041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1041に挿入される。シリアルポートインタフェース1050は、図12に例示するように、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、図12に例示するように、例えばディスプレイ1130に接続される。

[0073] ここで、図12に例示するように、ハードディスクドライブ1090は、例えば、OS (Operating System) 1091、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、上記のプログラムは、コンピュータ1000によって実行される指令が記述されたプログラムモジュールとして、例えばハードディスクドライブ1090に記憶される。

[0074] また、上記実施形態で説明した各種データは、プログラムデータとして、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、CPU 1020が、メモリ1010やハードディスクドライブ1090

0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、各種処理手順を実行する。

- [0075] なお、プログラムに係る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れず、例えば着脱可能な記憶媒体に記憶され、ディスクドライブ等を介してCPU1020によって読み出されてもよい。あるいは、プログラムに係るプログラムモジュール1093やプログラムデータ1094は、ネットワーク（LAN（Local Area Network）、WAN（Wide Area Network）等）を介して接続された他のコンピュータに記憶され、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。

符号の説明

- [0076] 1, 1A ソフトウェア解析システム
10, 10A 記憶部
11, 11A アプリ情報記憶部
12 キーワードリスト記憶部
20 ラベル生成部
30 スコア算出部
40 分類部
50 特定部
60 送信部

請求の範囲

【請求項1】 複数のアプリケーションのコードに基づき、当該複数のアプリケーションがパーミッションにより所定のプライバシー情報にアクセスするか否かを判定し、所定のプライバシー情報にアクセスすると判定されたアプリケーションに第1のラベルを付与し、所定のプライバシー情報にアクセスしないと判定されたアプリケーションに前記第1のラベルと異なる第2のラベルを付与するラベル生成部と、

前記第2のラベルを付与されたアプリケーションの説明文のテキストよりも、前記第1のラベルを付与されたアプリケーションの説明文のテキストに多く含まれる単語のスコアが高くなるよう前記複数のアプリケーションの説明文に含まれる単語各々のスコアを算出し、前記パーミッションに対応付けて、算出されたスコアが上位から所定数の単語を、前記複数のアプリケーションの説明文から抽出するスコア算出部と、

前記スコア算出部が抽出した単語が説明文のテキストに含まれるアプリケーションを前記パーミッションに言及するアプリケーションとして分類し、前記スコア算出部が抽出した単語が説明文のテキストに含まれないアプリケーションを前記パーミッションに言及しないアプリケーションとして分類する分類部と、

を備えるソフトウェア解析システム。

【請求項2】 前記ラベル生成部により前記第1のラベルを付与されたアプリケーションであって、前記分類部により前記パーミッションに言及しないアプリケーションに分類されたアプリケーションを、悪性アプリケーションとして特定する特定部を、さらに備える請求項1に記載のソフトウェア解析システム。

【請求項3】 前記ラベル生成部により前記第1のラベルを付与されたアプリケーションであって、前記分類部により前記パーミッションに言及しないアプリケーションに分類されたアプリケーションが、モバイル端末に

インストールされる際に、警告情報を送信する送信部をさらに備える請求項 1 または 2 に記載のソフトウェア解析システム。

[請求項4] 前記スコア算出部が抽出した単語と、前記パーミッションと、を対応づけて記憶する記憶部をさらに備え、

前記分類部は、新しいアプリケーションが検知されると、前記記憶部に前記パーミッションに対応付けて記憶される単語にもとづき、当該アプリケーションを、前記パーミッションに言及するアプリケーションおよび前記パーミッションに言及しないアプリケーションのいずれかに分類する請求項 1 から 3 のいずれか 1 項に記載のソフトウェア解析システム。

[請求項5] 前記スコア算出部は、少なくとも、前記第 1 のラベルを付与されたアプリケーションの総数、説明文のテキストが所定の単語を含むアプリケーションの総数、前記第 1 のラベルを付与されたアプリケーションのうち前記説明文のテキストが前記所定の単語を含むアプリケーションの総数に基づき、前記所定の単語のスコアを算出する請求項 1 から 4 のいずれか 1 項に記載のソフトウェア解析システム。

[請求項6] 前記スコア算出部は、式 (1) に基づいて、前記単語のスコアを算出する請求項 1 から 5 のいずれか 1 項に記載のソフトウェア解析システム。

[数1]

$$RW(w_i) = \log \frac{(r_i + 0.5)(N - n_i - R + r_i + 0.5)}{(n_i - r_i + 0.5)(R - r_i + 0.5)} \quad \dots (1)$$

(式 (1) 中、 w_i はアプリケーション A_i (i は 1 以上 N 以下の自然数) に含まれる単語であり、 R は前記第 1 のラベルを付与されたアプリケーションの総数であり、 r_i は前記複数のアプリケーションのうち単語 w_i を含むアプリケーションの総数であり、 N は前記複数のアプリケーションの総数であり、 n_i は単語 w_i を含むアプリケーションの総数である。)

[請求項7] 複数のアプリケーションのコードに基づき、当該複数のアプリケーションがパーミッションにより所定のプライバシー情報にアクセスするか否かを判定し、所定のプライバシー情報にアクセスすると判定されたアプリケーションに第1のラベルを付与し、所定のプライバシー情報にアクセスしないと判定されたアプリケーションに前記第1のラベルと異なる第2のラベルを付与するラベル生成工程と、

前記第2のラベルを付与されたアプリケーションの説明文のテキストよりも、前記第1のラベルを付与されたアプリケーションの説明文のテキストに多く含まれる単語のスコアが高くなるよう前記複数のアプリケーションの説明文に含まれる単語各々のスコアを算出し、前記パーミッションに対応付けて、算出されたスコアが上位から所定数の単語を、前記複数のアプリケーションの説明文から抽出するスコア算出工程と、

前記スコア算出工程において抽出した単語が説明文のテキストに含まれるアプリケーションを前記パーミッションに言及するアプリケーションとして分類し、前記スコア算出工程において抽出した単語が説明文のテキストに含まれないアプリケーションを前記パーミッションに言及しないアプリケーションとして分類する分類工程と、

を備えるソフトウェア解析方法。

[請求項8] 複数のアプリケーションのコードに基づき、当該複数のアプリケーションがパーミッションにより所定のプライバシー情報にアクセスするか否かを判定し、所定のプライバシー情報にアクセスすると判定されたアプリケーションに第1のラベルを付与し、所定のプライバシー情報にアクセスしないと判定されたアプリケーションに前記第1のラベルと異なる第2のラベルを付与するラベル生成手順と、

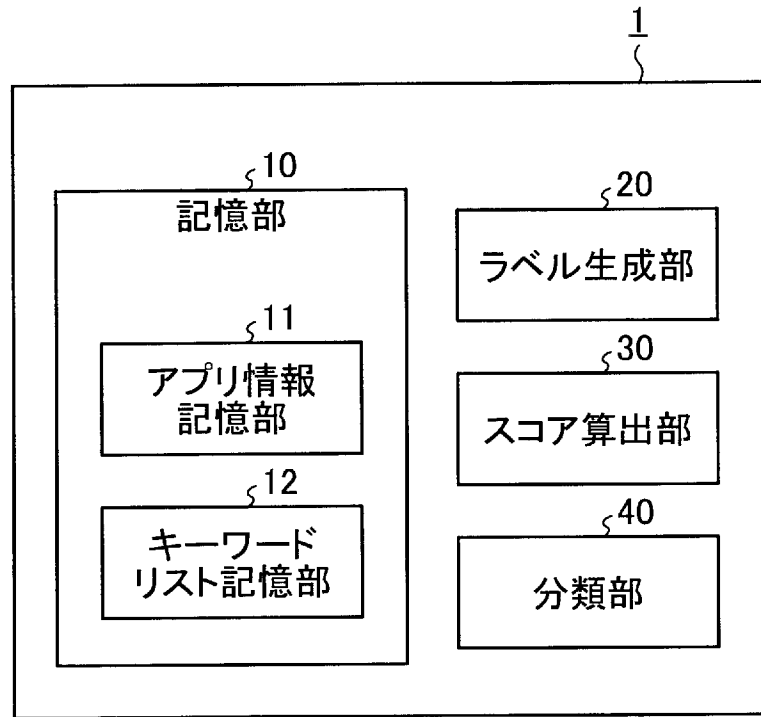
前記第2のラベルを付与されたアプリケーションの説明文のテキストよりも、前記第1のラベルを付与されたアプリケーションの説明文のテキストに多く含まれる単語のスコアが高くなるよう前記複数のア

アプリケーションの説明文に含まれる単語各々のスコアを算出し、前記パーミッションに対応付けて、算出されたスコアが上位から所定数の単語を、前記複数のアプリケーションの説明文から抽出するスコア算出手順と、

前記スコア算出手順において抽出した単語が説明文のテキストに含まれるアプリケーションを前記パーミッションに言及するアプリケーションとして分類し、前記スコア算出手順において抽出した単語が説明文のテキストに含まれないアプリケーションを前記パーミッションに言及しないアプリケーションとして分類する分類手順と、

をコンピュータに実行させるためのソフトウェア解析プログラム。

[図1]



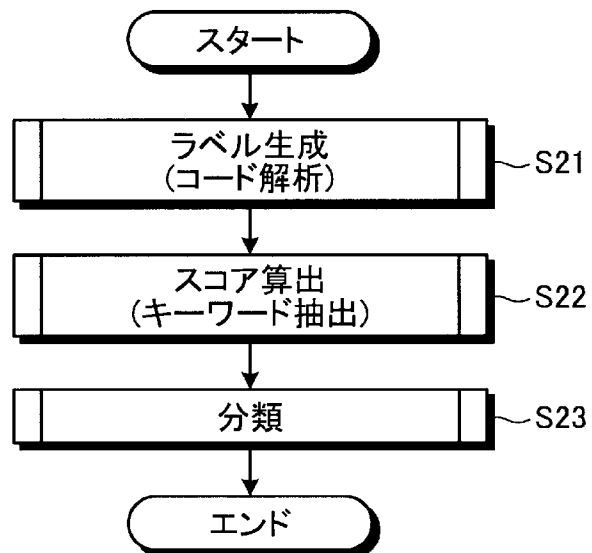
[図2]

アプリID	コード	説明文 テキスト	プライバシー 情報	ラベル	分類結果
AP001	~~~	~~~	カレンダー	言及あり	言及あり
			アドレス帳	言及なし	言及なし
			GPS情報	言及あり	言及なし
AP002	△△△	〇〇〇〇			

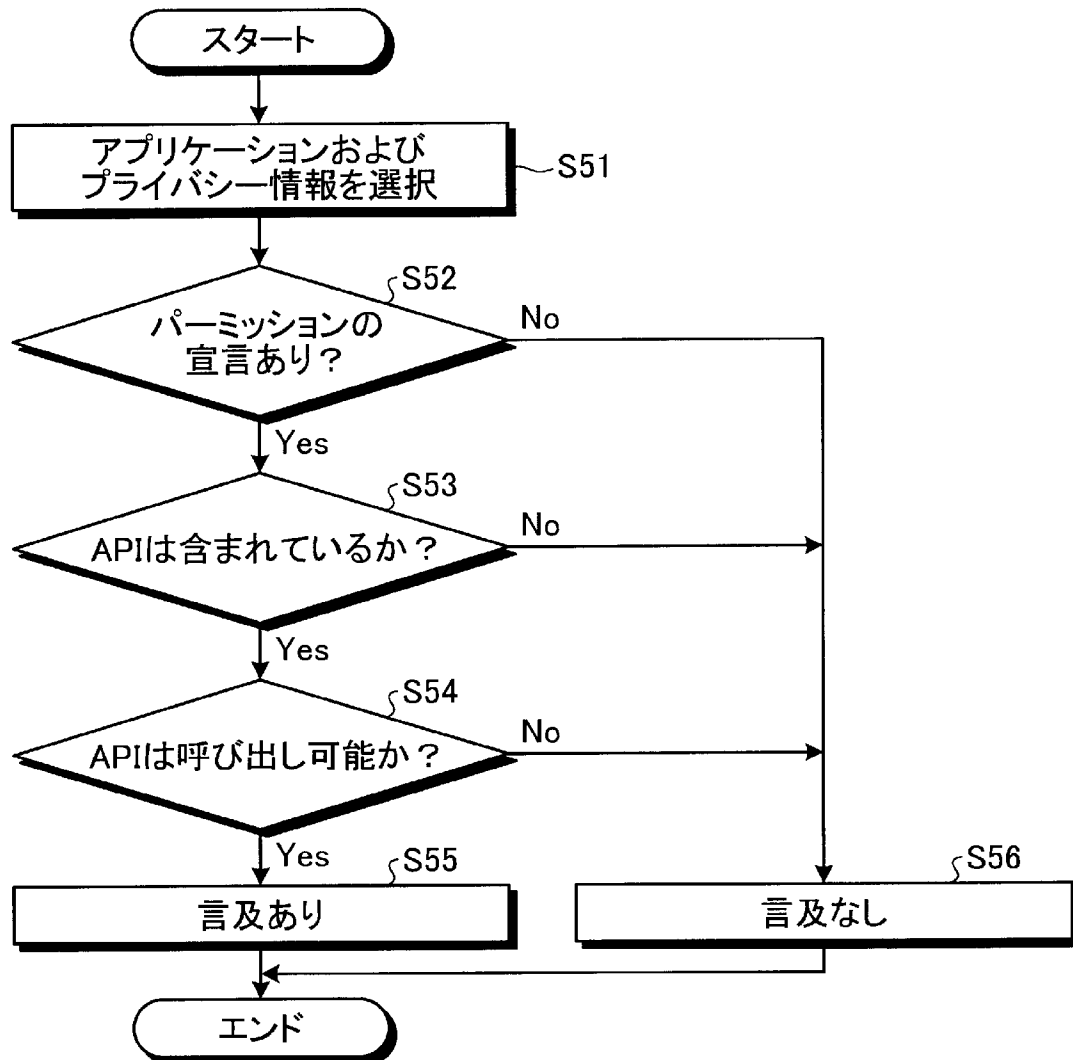
[図3]

プライバシー情報	機能 (パーミッション)	キーワード
GPS情報	Location	gps, location, map
アカウント情報	Account	username, account, password
連絡先	Contact	sms, call, contact
カレンダー	Calendar	calendar, reminder, meeting
受信SMSメール	SMS (read)	sms, message, incoming
送信SMSメール	SMS (sent)	sms, message, sent
カメラ機能	Camera	camera, scan, photo

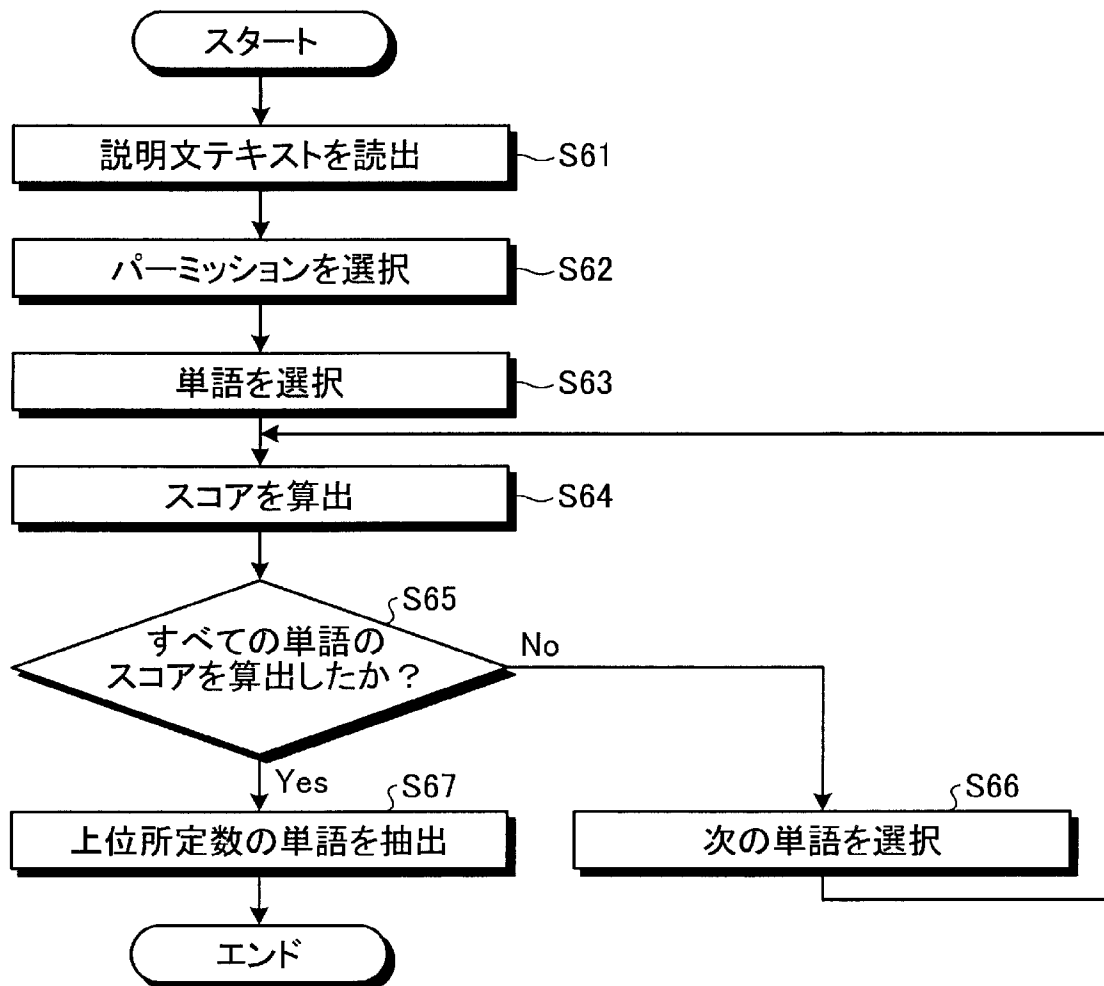
[図4]



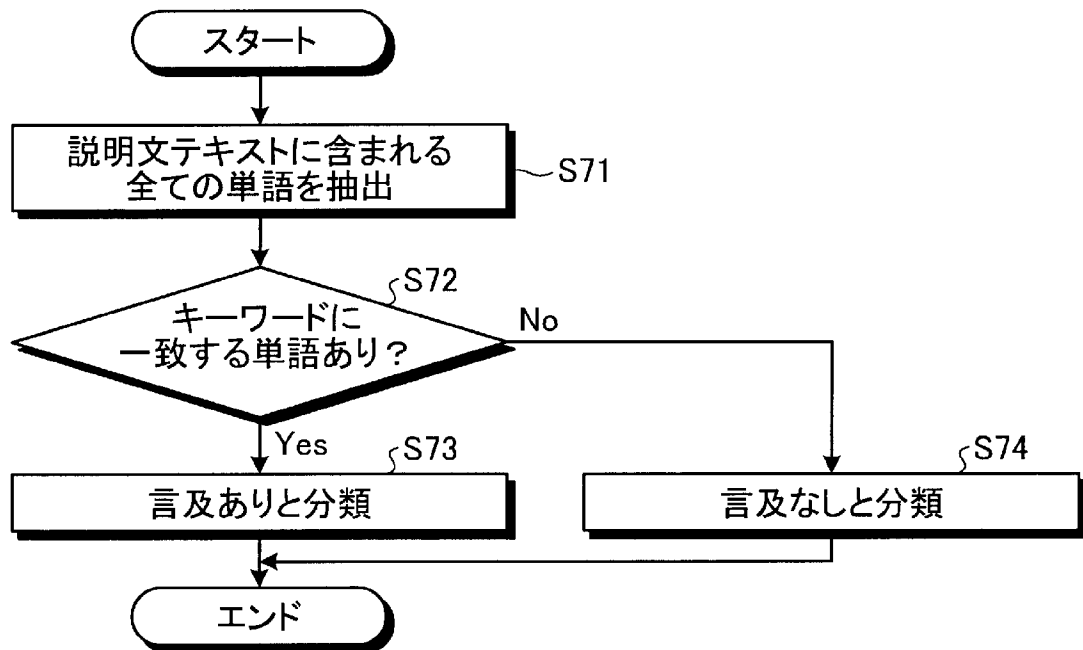
[図5]



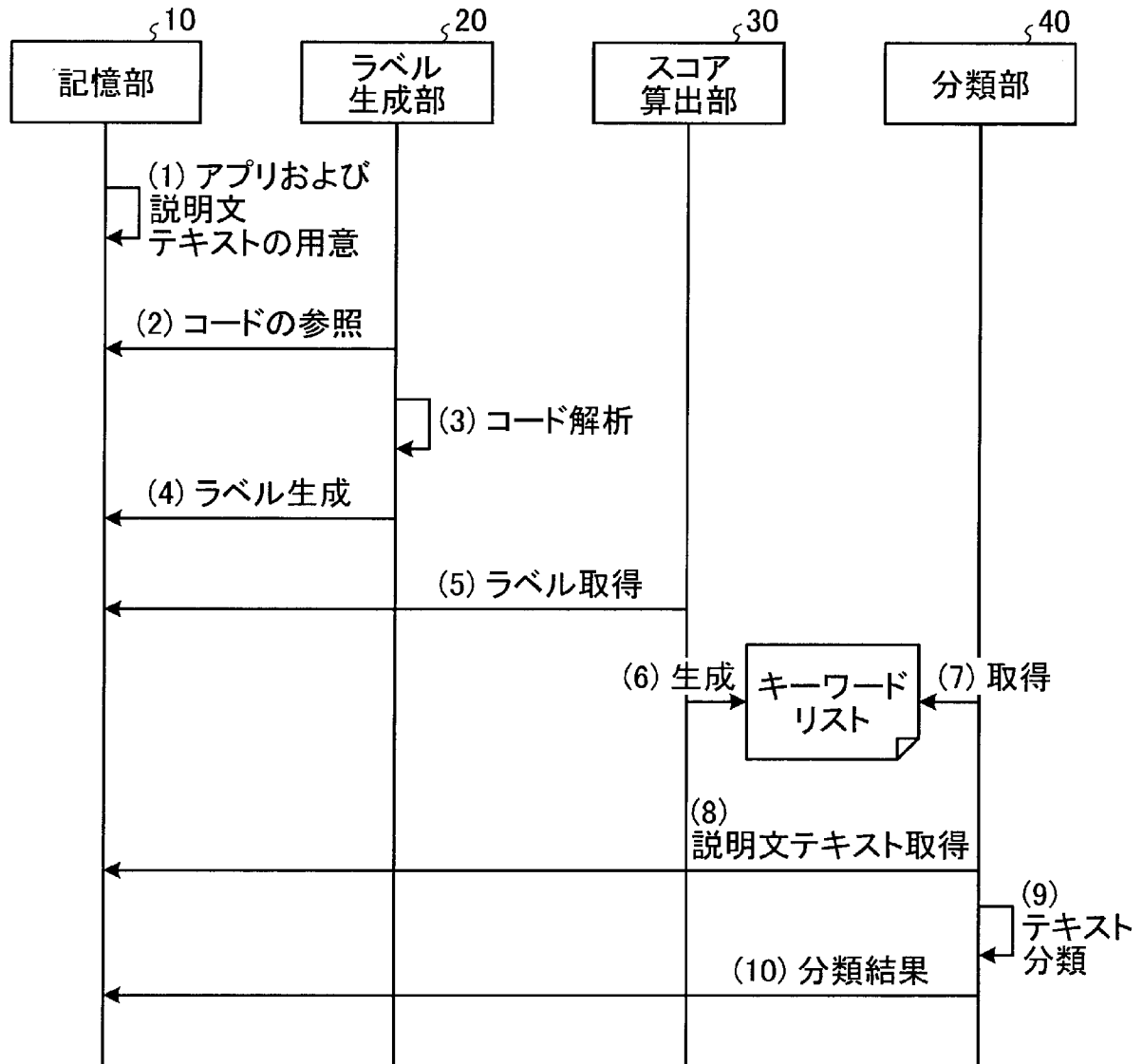
[図6]



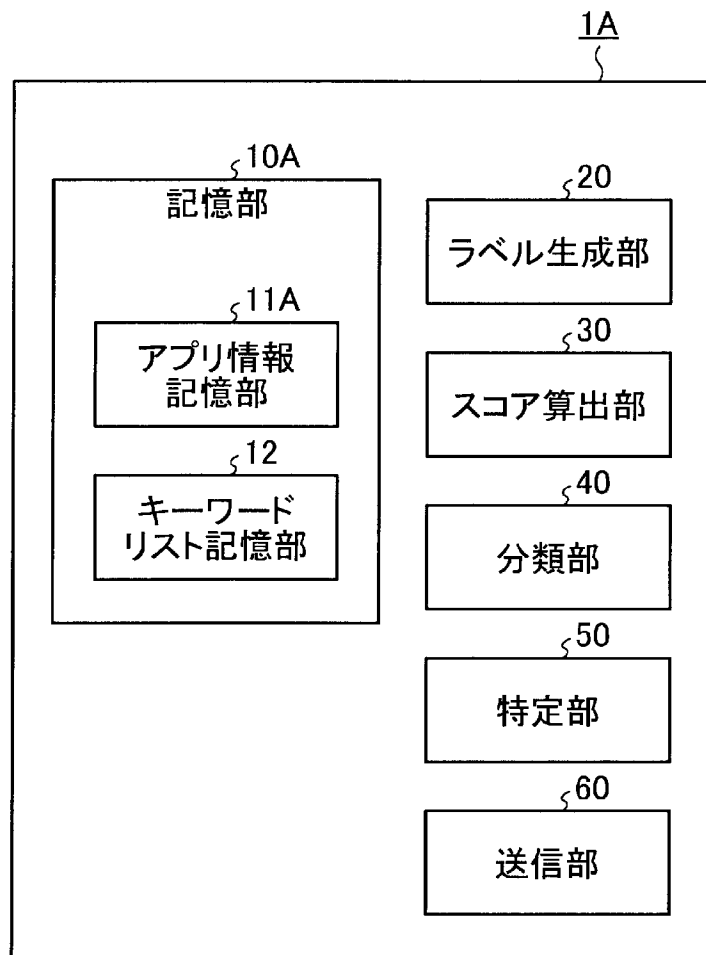
[図7]



[図8]



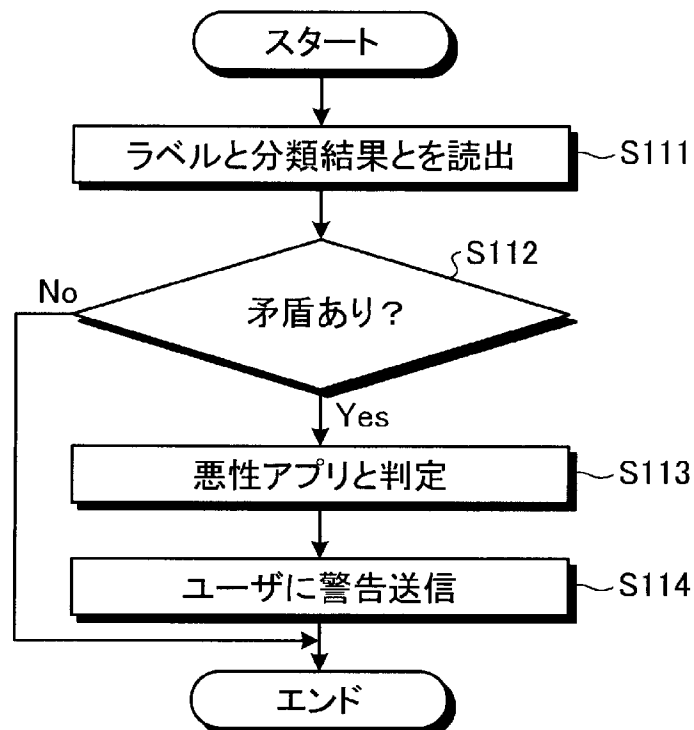
[図9]



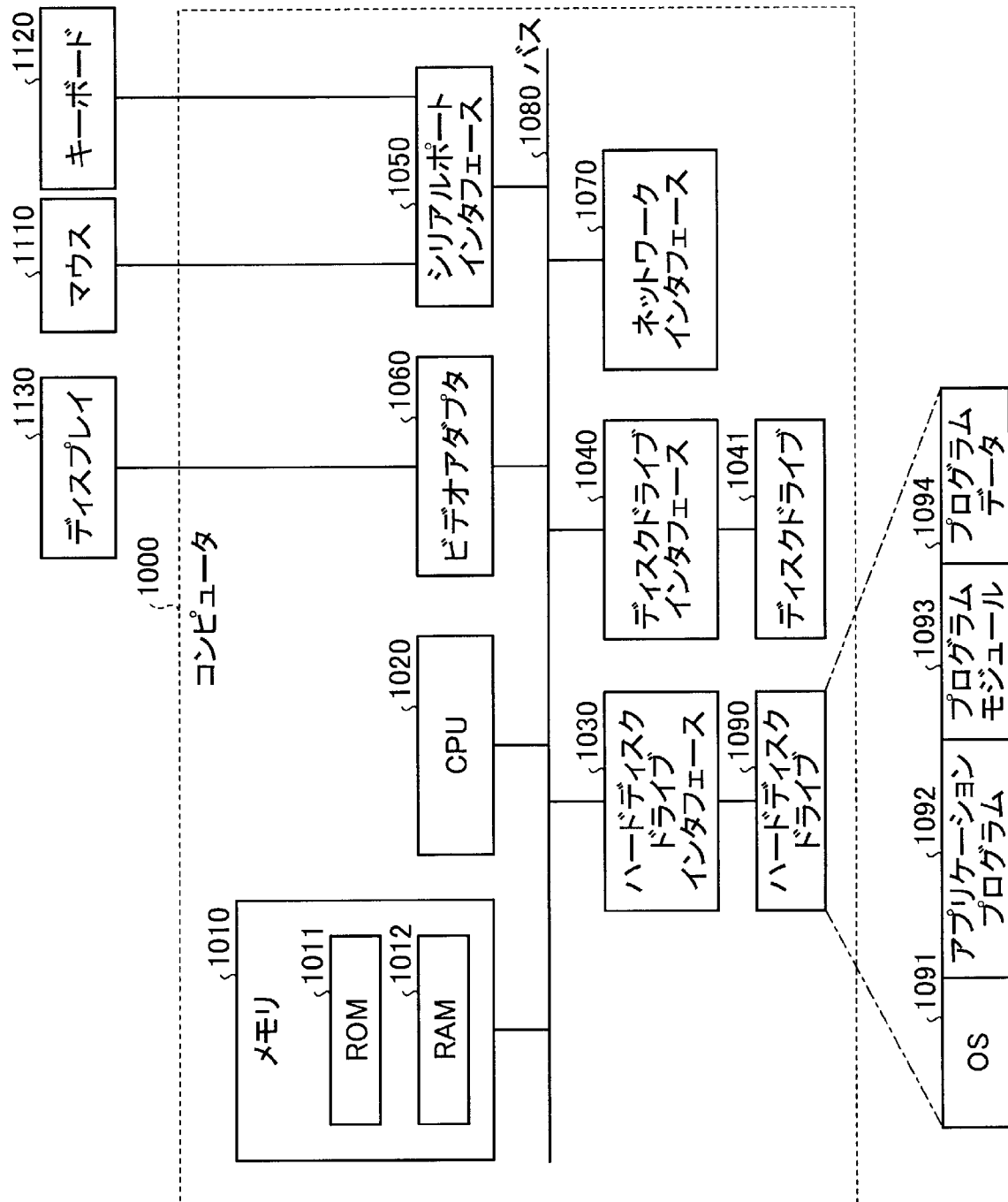
[図10]

アプリID	コード	説明文 テキスト	プライバシー 情報	ラベル	分類結果	フラグ
AP001	~~~	~~~	カレンダー	言及あり	言及あり	
			アドレス帳	言及なし	言及なし	
			GPS情報	言及あり	言及なし	ON
AP002	△△△	〇〇〇〇				

[図11]



[図12]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2016/069881

A. CLASSIFICATION OF SUBJECT MATTER

G06F21/62(2013.01) 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)

G06F21/62, G06F21/56

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	1996-2016
Kokai Jitsuyo Shinan Koho	1971-2016	Toroku Jitsuyo Shinan Koho	1994-2016

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	Takuya WATANABE et al., "Analyzing the Inconsistency between Words and Actions of Android Apps", CSS2014 Computer Security Symposium 2014 Ronbunshu IPSJ Symposium Series, vol.2014, no.2 [CD-ROM], 2014.10, vol.2014, pages 590 to 597	1-8
A	US 2013/0333039 A1 (KELLY), 12 December 2013 (12.12.2013), abstract; fig. 1 & WO 2013/184799 A1 & CN 104380302 A	1-8

☐ 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
29 September 2016 (29.09.16)

Date of mailing of the international search report
11 October 2016 (11.10.16)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G06F21/62(2013.01)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. G06F21/62, 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	1971-2016年
日本国実用新案登録公報	1996-2016年
日本国登録実用新案公報	1994-2016年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	渡邊卓弥他, Androidアプリの説明文とプライバシー情報アクセスの相関分析, CSS2014 コンピュータセキュリティシンポジウム2014 論文集 情報処理学会シンポジウムシリーズ Vol. 2014 No. 2 [CD-ROM], 2014.10, 第2014巻, p.590-p.597	1-8
A	US 2013/0333039 A1 (KELLY) 2013.12.12, ABSTRACT, FIG.1 & WO 2013/184799 A1 & CN 104380302 A	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

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

29.09.2016

国際調査報告の発送日

11.10.2016

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）

宮司 卓佳

5 S

9 5 5 5

電話番号 03-3581-1101 内線 3546