(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2020年9月3日(03.09.2020)

WIPO | PCT

(10) 国際公開番号
WO 2020/175113 A1

(51) 国際特許分類:
G06F 11/07 (2006.01)

(21) 国際出願番号： PCT/JP2020/004969

(22) 国際出願日： 2020年2月7日(07.02.2020)

(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 2019-033254 2019年2月26日(26.02.2019) 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).

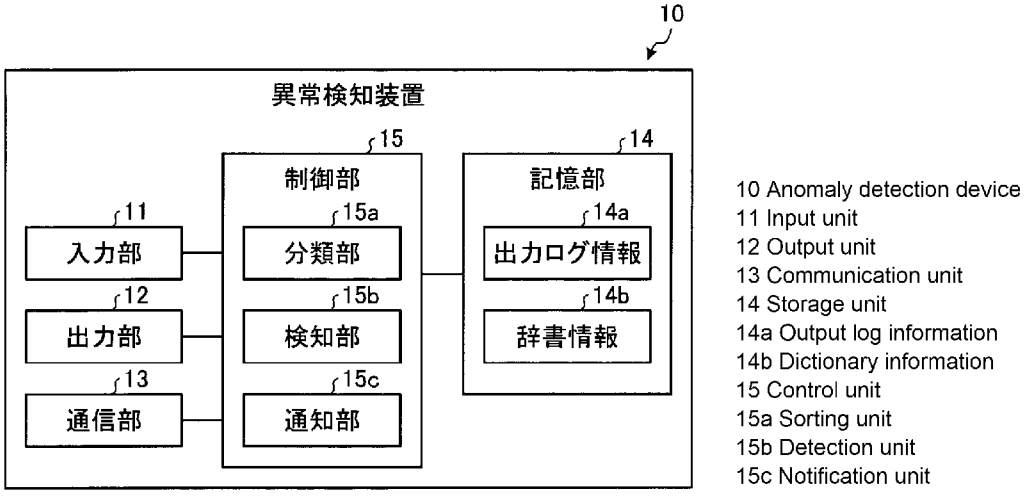
(72) 発明者: 東羅 翔太郎 (**TORA, Shotaro**); 〒1808585 東京都武蔵野市緑町３丁目９－１１ ＮＴＴ 知的財産センタ内 Tokyo (JP). 外山 将司 (**TOYAMA, Masashi**); 〒1808585 東京都武蔵野市緑町３丁目９－１１ ＮＴＴ 知的財産センタ内 Tokyo (JP). 豊田 真智子(**TOYODA, Machiko**); 〒1808585 東京都武蔵野市緑町３丁目９－１１ ＮＴＴ 知的財産センタ内 Tokyo (JP).

(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).

(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,

(54) **Title:** ANOMALY DETECTION DEVICE, ANOMALY DETECTION METHOD, AND ANOMALY DETECTION PROGRAM

(54) 発明の名称： 異常検知装置、異常検知方法および異常検知プログラム



(57) **Abstract:** An anomaly detection device (10) is provided with a storage unit (14) that stores dictionary information (14b) in which partial message text strings that represent the types of messages included in a text log output from a system are respectively associated with an ID that is set for each message type. When the anomaly detection device (10) acquires messages included in a text log output from the system, the anomaly detection device (10) refers to the dictionary information (14b) stored in the storage unit (14), sorts the messages included in the text log by type, applies an ID to each sorted message, and detects anomalies on the basis of the IDs applied to the messages.

WO 2020/175113 A1

[続葉有]

ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))

(57) 要約:異常検知装置(１０)は、システムから出力されたテキストログに含まれる各メッセージの種別を表す各メッセージの部分文字列と、メッセージの種別ごとに設定されたＩＤとが対応付けられた辞書情報(１４ｂ)を記憶する記憶部(１４)を有する。異常検知装置(１０)は、システムから出力されたテキストログに含まれるメッセージを取得すると、記憶部(１４)に記憶された辞書情報(１４ｂ)を参照し、テキストログに含まれるメッセージを種別ごとに分類し、分類した各メッセージにＩＤを付与し、メッセージに付与されたＩＤに基づいて、異常を検知する。

明　細　書

発明の名称 ：

異常検知装置、異常検知方法および異常検知プログラム

技術分野

[0001] 本発明は、異常検知装置、異常検知方法および異常検知プログラムに関する。

背景技術

[0002] 従来、サーバシステムやネットワークシステムにおける異常検知や状態分析のために、ｓｙｓｌｏｇやＭＩＢ（Management Information Base）情報等のテキストログを用いたシステムの監視が行われている。

[0003] 例えば、システムに障害が発生した際に、人手によって特定のキーワードでテキストログを検索し、当該キーワードを含んだメッセージをクリティカルなメッセージとして抽出することが行われている。

先行技術文献

特許文献

[0004] 特許文献1：特開２０１４－１５３７２３号公報

特許文献2：特開２０１５－３６８９１号公報

発明の概要

発明が解決しようとする課題

[0005] しかしながら、従来のブラックリストによる監視では、既知の異常を検知することはできても、未知の異常を検知することはできないという課題があった。

課題を解決するための手段

[0006] 上述した課題を解決し、目的を達成するために、本発明の異常検知装置は、システムから出力されたテキストログに含まれる各メッセージの種別を表す各メッセージの部分文字列と、メッセージの種別ごとに設定されたＩＤと

が対応付けられた辞書情報を記憶する記憶部と、前記システムから出力されたテキストログに含まれるメッセージを取得すると、前記記憶部に記憶された辞書情報を参照し、前記テキストログに含まれるメッセージを種別ごとに分類し、分類した各メッセージにＩＤを付与する分類部と、前記分類部によってメッセージに付与されたＩＤに基づいて、異常を検知する検知部とを有することを特徴とする。

[0007] また、本発明の異常検知方法は、異常検知装置によって実行される異常検知方法であって、前記異常検知装置は、システムから出力されたテキストログに含まれる各メッセージの種別を表す各メッセージの部分文字列と、メッセージの種別ごとに設定されたＩＤとが対応付けられた辞書情報を記憶する記憶部を有し、前記システムから出力されたテキストログに含まれるメッセージを取得すると、前記記憶部に記憶された辞書情報を参照し、前記テキストログに含まれるメッセージを種別ごとに分類し、分類した各メッセージにＩＤを付与する分類工程と、前記分類工程によってメッセージに付与されたＩＤに基づいて、異常を検知する検知工程とを含むことを特徴とする。

[0008] また、本発明の異常検知プログラムは、システムから出力されたテキストログに含まれるメッセージを取得すると、前記システムから出力されたテキストログに含まれる各メッセージの種別を表す各メッセージの部分文字列とメッセージの種別ごとに設定されたＩＤとが対応付けられた辞書情報を記憶する記憶部に記憶された前記辞書情報を参照し、前記テキストログに含まれるメッセージを種別ごとに分類し、分類した各メッセージにＩＤを付与する分類ステップと、前記分類ステップによってメッセージに付与されたＩＤに基づいて、異常を検知する検知ステップとをコンピュータに実行させることを特徴とする。

発明の効果

[0009] 本発明によれば、未知の異常を検知することができるという効果を奏する。

図面の簡単な説明

[0010] [図1]図１は、第１の実施形態に係る異常検知装置の構成例を示すブロック図である。

[図2]図２は、テキストログの一例を示す図である。

[図3]図３は、辞書情報のデータ構成の一例を示す図である。

[図4]図４は、テンプレートの作成について説明するための図である。

[図5]図５は、ＩＤ付与処理を説明する図である。

[図6]図６は、ログを集約してＩＤを付与する一連の流れの一例を説明する図である。

[図7]図７は、システムが安定して運用されている場合の新規テンプレート数の推移を示す図である。

[図8]図８は、未知の事象が発生している場合の新規テンプレート数の推移を示す図である。

[図9]図９は、単位時間当たりの新規テンプレート数を監視することで未知異常を発見する一連の処理の流れを説明する図である。

[図10]図１０は、ＩＤごとの出現頻度の相対的な変化から異常度を算出する処理を説明する図である。

[図11]図１１は、第１の実施形態に係る異常検知装置におけ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、異常検知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

[0011] 以下に、本願に係る異常検知装置、異常検知方法および異常検知プログラムの実施の形態を図面に基づいて詳細に説明する。なお、この実施の形態により本願に係る異常検知装置、異常検知方法および異常検知プログラムが限定されるものではない。

[0012] ［第１の実施形態］

以下の実施の形態では、第１の実施形態に係る異常検知装置１０の構成、異常検知装置１０の処理の流れを順に説明し、最後に第１の実施形態による

効果を説明する。

[0013] ［異常検知装置の構成］

まず、図１を用いて、本実施形態の異常検知装置１０の構成例を説明する。図１は、第１の実施形態に係る異常検知装置の構成例を示すブロック図である。図１に示すように、異常検知装置１０は、入力部１１、出力部１２、通信部１３、記憶部１４及び制御部１５を有する。

[0014] 入力部１１は、ユーザからのデータの入力を受け付ける。入力部１１は、例えば、マウスやキーボード等の入力装置である。出力部１２は、画面の表示等により、データを出力する。出力部１２は、例えば、ディスプレイ等の表示装置である。通信部１３は、ネットワークを介して、他の装置との間でデータ通信を行う。例えば、通信部１３はＮＩＣ（Network Interface Card）である。

[0015] 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、ＮＶＳＲＡＭ（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、異常検知装置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。

[0016] また、記憶部１４は、出力ログ情報１４１及び辞書情報１４２を記憶する。記憶部１４は、システムから出力されたテキストログに含まれる各メッセージの種別を表す各メッセージの部分文字列と、メッセージの種別ごとに設定されたＩＤとが対応付けられた辞書情報１４２を記憶する。

[0017] 記憶部１４は、システムから出力されたテキストログを出力ログ情報１４ａとして記憶する。ここで、テキストログは、例えば、計算機システムを構成するサーバマシン、パーソナルコンピュータ、ストレージ等から出力される。また、テキストログは、例えば、ネットワークシステムを構成するルータ、ファイアウォール、ロードバランサ、光伝送装置、光伝送中継器等から

出力される。また、出力されるテキストログは、システム全体に関するものであってもよいし、システムを構成する装置に関するものであってもよい。さらに、テキストログは、計算機システムやネットワークシステムが仮想化された環境において出力されたものであってもよい。

[0018] テキストログは、例えば、ＯＳのシスログ、アプリケーション及びデータベースの実行ログ、エラーログ、操作ログ、ネットワーク機器から得られるＭＩＢ情報、監視システムのアラート、行動ログ、動作状態ログ等である。なお、テキストログは上記のものに限定されるものではなく、どのようなものであってもよい。

[0019] 図２は、テキストログの一例を示す図である。図２に示すように、テキストログの各レコードは、メッセージとメッセージに付された発生日時とを含む。例えば、テキストログの１行目のレコードは、メッセージ「LINK-UP Interface 1/0/17」と、発生日時「2015/05/18T14:56」とを含む。なお、メッセージには、ホストやログレベルといった付加情報を含んでも構わない。

[0020] 記憶部１４は、テキストログのメッセージを分類するためのデータを辞書情報１４ｂとして記憶する。図３は、辞書情報のデータ構成の一例を示す図である。図３に示すように、辞書情報１４ｂは、ＩＤ及び単語シーケンスで構成されるテンプレートを含む。つまり、テンプレートは、テンプレートＩＤ、単語シーケンスから構成される。ＩＤは、テキストログのメッセージが分類される種別を識別するための情報である。単語シーケンスは、テキストログのメッセージの分類に用いられる文字列である。例えば、ＩＤが「１」である種別にメッセージが分類されるか否かは、単語シーケンス「LINK-UP Interface ＊」を用いて判定される。なお、辞書情報１４ｂを用いたメッセージの分類は分類部１５ａによって行われる。分類部１５ａの具体的な処理については後述する。

[0021] 制御部１５は、異常検知装置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 Circu

it）、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。例えば、制御部１５は、分類部１５ａ、検知部１５ｂおよび通知部１５ｃを有する。

[0022] 分類部１５ａは、システムから出力されたテキストログに含まれるメッセージを取得すると、記憶部１４に記憶された辞書情報を参照し、テキストログに含まれるメッセージを種別ごとに分類し、分類した各メッセージにＩＤを付与する。前述の通り、分類部１５ａは、辞書情報１４ｂを用いて分類を行う。

[0023] ここで、辞書情報１４ｂは、分類部１５ａによって作成されてもよい。例えば、分類部１５ａは、メッセージからパラメータを削除することで得られる文字列を基にテンプレートを作成することができる。ここで、図４を用いて、テキストログに基づいたテンプレートの作成方法について説明する。図４は、テンプレートの作成について説明するための図である。テンプレートは、テンプレートＩＤ、単語シーケンスから構成される。

[0024] 分類部１５ａは、メッセージを所定の区切り文字で分割して単語群を抽出し、各単語を文字種によってパラメータまたは非パラメータに分類し、記憶部１４に記憶された辞書情報のテンプレート群のそれぞれの単語シーケンスと比較し、上記メッセージ中の非パラメータと分類された箇所の単語がすべて一致するテンプレートが存在する場合には、該テンプレートのＩＤをメッセージに付与する。

[0025] また、分類部１５ａは、メッセージの分類を行う際に、記憶部１４に記憶された辞書情報のテンプレート群の中に、非パラメータと分類された箇所の単語がすべて一致する単語シーケンスを持つテンプレートが存在しない場合には、当該メッセージを基にした単語シーケンスを持つ新たなテンプレートを作成し、新規ＩＤを付与した新しいテンプレートを生成しても良い。

[0026] 図４に示すように、分類部１５ａは、例えば、「数値と記号」で構成される単語をパラメータとみなし、メッセージからパラメータを削除した文字列を単語シーケンスとすることができる。この場合、分類部１５ａは、図２のテキストログに含まれるメッセージ「LINK-UP Interface 1/0/17」、又はメッセージ「LINK-UP Interface 0/0/0」から、「LINK-UP Interface」という単語シーケンスを作成する。さらに、分類部１５ａは、作成した単語シーケンスとＩＤとを持つテンプレートとして生成し、記憶部１４の辞書情報１４ｂに追加する。このとき、分類部１５ａは、パラメータを削除した部分に「*」等のワイルドカードを追加してもよい。

[0027] なお、分類部１５ａがパラメータとみなす単語は上記の例に限られない。分類部１５ａは、例えば、「数字を含む単語」を全てパラメータとみなしてもよいし、「数字とアルファベットで構成される単語」は非パラメータとみなしてもよいし、「アルファベットのみで構成される単語」だけを非パラメータとしてもよい。

[0028] また、パラメータ／非パラメータの判断基準は文字種に限らずルールベースで作成しても良い。具体的には、「IPアドレスのルールに従う単語」をパラメータとみなしてもよい。さらに、単語の枠を超えて、前記「所定の区切り文字」を含む文字列についてパラメータ／非パラメータと判定しても良い。

[0029] また、本実施形態において、辞書情報１４ｂのテンプレートは、分類部１５ａによって作成されたものである必要はなく、あらかじめユーザにより作成されたものや、異常検知装置１０以外の装置によって自動的に作成されたものであってもよい。

[0030] ここで、図５を用いて、分類部１５ａによるＩＤ付与処理を説明する。図５はG、ＩＤ付与処理を説明する図である。分類部１５ａは、辞書情報１４ｂ中の各テンプレートの単語シーケンスとテキストログのメッセージと比較することにより、メッセージの分類を行う。このとき、分類部１５ａは、単語シーケンスがメッセージに完全一致する場合、又は、部分一致する場合に、

テンプレートがメッセージに一致すると判定する。そして、分類部１５ａは、メッセージに対し、一致すると判定されたテンプレートのＩＤを付与する。

[0031] 例えば、図３のＩＤが「１」であるテンプレートの単語シーケンス「LINK-UP Interface *」が、図２のテキストログの１行目のメッセージ「LINK-UP Interface 1/0/17」に部分一致するため、分類部１５ａはテンプレートの単語シーケンス「LINK-UP Interface *」がメッセージ「LINK-UP Interface 1/0/17」に一致すると判定し、メッセージ「LINK-UP Interface 1/0/17」にＩＤ「１」を付与する。

[0032] また、分類部１５ａは、メッセージに一致する単語シーケンスを持つテンプレートが辞書情報１４ｂに存在しない場合には、まだ付与されていない新規ＩＤを付与するとともに、メッセージをもとに新たなテンプレートを辞書情報１４ｂに追加する。

[0033] 分類部１５ａは、メッセージの分類及びＩＤの付与を行うことで、分類済みテキストログを作成する。図２に示すように、分類済みテキストログの各レコードは、メッセージのＩＤ及び発生日時を含む。例えば、分類済みテキストログの１行目のレコードは、メッセージ「LINK-UP Interface 1/0/17」に付与されたＩＤ「１」と、発生日時「2015/05/18T14:56」とを含む。

[0034] 次に、図６を用いて、ログを集約してＩＤを付与する一連の流れの一例を説明する。図６は、ログを集約してＩＤを付与する一連の流れの一例を説明する図である。図６に例示するように、分類部１５ａは、ファイルログやリアルタイムログ等のフォーマットの異なる種々のログをそれぞれの形式に従ってフォーマット変換する。ここで、例えば、分類部１５ａは、リアルタイムログについて、時刻の入れ替わりを防ぐため、ログを受け取った時刻を付与した後に、フォーマットを変換する。フォーマット変換済みログメッセージは、日時やメッセージ本文、その他の情報を含むものとする。

[0035] そして、分類部１５ａは、辞書情報のテンプレートを参照してメッセージにＩＤを付与し、ＩＤ化されたログメッセージを出力する。また、分類部１

５ａは、新規のＩＤを付与した場合には、新しいテンプレートを追加するように辞書情報を更新する。

[0036] 図１の説明に戻って、検知部１５ｂは、分類部１５ａによってメッセージに付与されたＩＤに基づいて、異常を検知する。例えば、検知部１５ｂは、分類部１５ａによって分類されたメッセージのうち、記憶部１４の辞書情報の中に該当するメッセージの種別が存在しなかったメッセージの種別の件数が所定の閾値を超えている場合には、異常を検知する。

[0037] ここで、図７および図８を用いて、新規テンプレート数の推移について説明する。図７は、システムが安定して運用されている場合の新規テンプレート数の推移を示す図である。図８は、未知の事象が発生している場合の新規テンプレート数の推移を示す図である。図７に例示するように、システムが安定して運用されていれば、初期段階においてある程度の期間は未知のログが出現するが、未知のログの出現頻度は減少していく。そして、図８に例示するように、システムにおいて障害やメンテナンス等の未知の事象が発生している場合には、所定の期間において未知のログが大量に出現する。

[0038] 検知部１５ｂは、このような未知の事象が発生していることを検知するため、所定の期間における新規ＩＤの付与数が所定の閾値を超えるか判定する。例えば、図８の例では、検知部１５ｂは、１日ごとの新規ＩＤの付与数を計数し、１日の新規ＩＤの付与数が閾値２５０を超えているかを監視する。そして、検知部１５ｂは、１日の新規ＩＤの付与数が閾値「２５０」を超えている場合に、異常を検知する。なお、閾値は、任意に設定変更可能であるものとする。

[0039] 図１の説明に戻って、通知部１５ｃは、分類部１５ａによって付与されたＩＤを含むメッセージに関する情報を通知する。例えば、通知部１５ｃは、新規ＩＤリストを外部の端末装置に通知するようにしてもよい。また、例えば、通知部１５ｃは、検知部１５ｂによって異常が検知された場合には、外部の端末装置にアラートを通知するようにしてもよい。また、例えば、通知部１５ｃは、外部の端末装置からの要求に応じて、辞書情報やＩＤのヒート

マップを表示するようにしてもよい。

[0040] ここで、図９を用いて、単位時間当たりの新規ＩＤ付与数を監視することで未知異常を発見する一連の処理の流れを説明する。図９は、単位時間当たりの新規ＩＤ付与数を監視することで未知異常を発見する一連の処理の流れを説明する図である。図９に例示するように、異常検知装置１０は、異常が発生したシステムから出力された新規ログについて、記憶部１４に記憶された辞書情報１４ｂを参照し、辞書情報に登録されていないテキストログのメッセージには新規ＩＤを付与する。そして、異常検知装置１０は、新規ＩＤリストを外部の端末装置に通知する。

[0041] また、異常検知装置１０では、時刻毎の新規ＩＤ数と辞書を常時可視化できるようにし、何かあったときに迅速に原因特定し、新たな監視項目にすることができる。また、異常検知装置１０は、所定の期間における新規ＩＤ数が所定の閾値を超えたときにはアラートを外部の端末装置に通知する。このように、異常検知装置１０では、単位時間あたりの新規ＩＤ付与数を監視することで未知異常を早期に発見することができ、ユーザ申告前のトラブル対処を行うことができる。

[0042] つまり、正常時のログが辞書情報１４ｂに登録されていれば異常時のログは新規ＩＤが付与されるという前提のもと、異常検知装置１０は、辞書情報にないログには新規ＩＤを付与し、新規ＩＤの増加数から未知の異常を検知することが可能である。

[0043] また、異常検知装置１０では、端末装置からの要求に応じて、辞書ビューアやＩＤのヒートマップを表示するようにしてもよく、辞書ビューアやログヒートマップとして可視化することで、迅速に内容を確認することが可能である。

[0044] また、上記の説明では、異常検知装置１０の検知部１５ｂが、所定の期間における新規ＩＤ数が所定の閾値を超えたときに異常を検知する場合を説明したが、これに限定されるものではない。例えば、検知部１５ｂは、分類部１５ａによってメッセージが分類された結果から、メッセージに付与された

ＩＤの単位時間（例えば、１日）あたりの付与回数をＩＤごとに集計し、任意の単位時間の付与回数と他の単位時間の付与回数とを比較し、各単位時間のＩＤごとの異常度を検知するようにしてもよい。

[0045] ここで、図１０を用いて、ＩＤごとの出現頻度の相対的な変化から異常度を算出する処理について説明する。図１０は、ＩＤごとの出現頻度の相対的な変化から異常度を算出する処理を説明する図である。例えば、検知部１５ｂは、毎日、１日ごとの集計値を過去５日分算出し、ＴＦ－ＩＤＦ値を計算すると、直近５日間と比べて１日あたりの頻度が多いＩＤを見付けることができる。図１０の左側においては、「１」～「７」のＩＤについて、５日間（４日前、３日前、２日前、昨日、今日）における１日ごとの付与数の和がそれぞれ表示されている。

[0046] 例えば、検知部１５ｂは、図１０に例示する式を用いて、各ＩＤの１日ごとのＴＦ－ＩＤＦ値を計算する。具体例を挙げて説明すると、検知部１５ｂは、例えば、ＩＤ「４」の「昨日」について、「$５\times \log_{10}(５／１)$」を計算し、ＴＦ－ＩＤＦ値として約「３．５」となる。

[0047] そして、検知部１５ｂは、計算したＴＦ－ＩＤＦ値を異常度とし、異常度が所定の閾値（例えば、０．７）以上であるＩＤが存在する場合には、異常を検知するようにしてもよい。

[0048] ［異常検知装置の処理手順］

次に、図１１を用いて、第１の実施形態に係る異常検知装置１０による処理手順の例を説明する。図１１は、第１の実施形態に係る異常検知装置における処理の流れの一例を示すシーケンス図である。なお、図１１の例では、新規ＩＤが付与されるたびに、新規ＩＤリストを通知し、異常を検知する場合の処理例を説明するが、新規ＩＤリストを通知する処理や異常を検知する処理を行うタイミングはこれに限定されるものではなく、どのようなタイミングであってもよい。

[0049] 図１１に例示するように、異常検知装置１０の分類部１５ａは、ログメッセージを受け付けると（ステップＳ１０１肯定）、メッセージを所定の区切

り文字で分割して単語群を抽出する（ステップＳ１０２）。そして、分割部１５ａは、各単語を文字種によってパラメータまたは非パラメータに分類し、記憶部１４に記憶された辞書情報のテンプレート群のそれぞれの単語シーケンスと比較し、上記メッセージ中の非パラメータと分類された箇所の単語がすべて一致するテンプレートが存在する場合には、該テンプレートのＩＤをメッセージに付与する（ステップＳ１０３）。

[0050] そして、分類部１５ａは、ステップＳ１０３において新規ＩＤを付与したか判定する（ステップＳ１０４）。この結果、分類部１５ａは、新規ＩＤを付与したと判定しなかった場合には（ステップＳ１０４否定）、そのまま処理を終了する。また、分類部１５ａは、新規ＩＤを付与したと判定した場合には（ステップＳ１０４肯定）、記憶部１４の辞書情報１４ｂに新しいテンプレートを追加する（ステップＳ１０５）。

[0051] 続いて、通知部１５ｃは、新規ＩＤリストを通知する（ステップＳ１０６）。そして、検知部１５ｂは、単位時間あたりの新規テンプレート数が閾値を超えるか判定する（ステップＳ１０７）。この結果、検知部１５ｂは、単位時間あたりの新規テンプレート数が所定数を超えていない場合には（ステップＳ１０７否定）、そのまま処理を終了する。また、検知部１５ｂは、単位時間あたりの新規テンプレート数が超えている場合には（ステップＳ１０７肯定）、異常を検知する（ステップＳ１０８）。

[0052] ［第１の実施形態の効果］

第１の実施形態に係る異常検知装置１０は、過去にシステムから出力されたテキストログに含まれるメッセージと、メッセージの種別ごとに設定されたＩＤとが対応付けられた辞書情報１４ｂを記憶する記憶部１４を有する。異常検知装置１０は、システムから出力されたテキストログに含まれるメッセージを取得すると、記憶部１４に記憶された辞書情報１４ｂを参照し、テキストログに含まれるメッセージを種別ごとに分類し、分類した各メッセージにＩＤを付与し、メッセージに付与されたＩＤに基づいて、異常を検知する。このため、異常検知装置１０では、未知の異常を検知することが可能で

ある。

[0053] また、第１の実施形態に係る異常検知装置１０は、メッセージに付与されたＩＤのうち、所定の期間における新規ＩＤの付与数が所定の閾値を超えている場合には、異常を検知するので、単位時間あたりの新規ＩＤ付与数を監視することで未知異常を早期に検知することが可能である。

[0054] また、第１の実施形態に係る異常検知装置１０は、ＩＤを含むメッセージに関する情報を通知することで、例えば、時刻ごとの新規ＩＤ数の変化をユーザが常時監視することが可能である。

[0055] また、第１の実施形態に係る異常検知装置１０は、前記メッセージに付与されたＩＤの単位時間あたりの頻度をＩＤごとに集計し、任意の単位時間の頻度と他の単位時間の頻度とを比較し、各単位時間のＩＤごとの異常度を検知することで、過去に観測されたことがあるＩＤのログについても異常を検知することが出来る。

[0056] また、第１の実施形態に係る異常検知装置１０は、分類部１５ａが、メッセージを所定の区切り文字で分割して単語群を抽出し、各単語を文字種によってパラメータまたは非パラメータに分類し、記憶部１４に記憶された辞書情報のテンプレート群のそれぞれの単語シーケンスと比較し、上記メッセージ中の非パラメータと分類された箇所の単語がすべて一致するテンプレートが存在する場合には、該テンプレートのＩＤをメッセージに付与することで、分類方法について事前に詳細な設定を行わずにシステムのログメッセージを分類することが可能である。

[0057]（システム構成等）

また、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。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。

[0058] また、本実施の形態において説明した各処理のうち、自動的におこな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おこなうこともでき、あるいは、手動的におこなわれるものとして説明した処理の全部また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。

[0059] （プログラム）

また、上記実施形態において説明した異常検知装置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、実施形態に係る異常検知装置１０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異常検知プログラムを作成することもできる。この場合、コンピュータが異常検知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、かかる異常検知プログラムをコンピュータ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異常検知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。

[0060] 図１２は、異常検知プログラムを実行するコンピュータを示す図である。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、ＣＰＵ１０２０と、ハードディスクドライブインタフェース１０３０と、ディスクドライブインタフェース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。

[0061] メモリ１０１０は、図１２に例示するように、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ブインタフェース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。ディスクドライブインタフェース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、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。

[0062] 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、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、上記の、異常検知プログラムは、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。

[0063] 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、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、各種処理手順を実行する。

[0064] なお、異常検知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、異常検知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

[0065] １０　異常検知装置

１１　入力部

１２　出力部

１３　通信部

１４　記憶部

１４ａ　出力ログ情報

１４ｂ　辞書情報

１５　制御部

１５ａ　分類部

１５ｂ　検知部

１５ｃ　通知部

請求の範囲

[請求項1] システムから出力されたテキストログに含まれる各メッセージの種別を表す各メッセージの部分文字列と、メッセージの種別ごとに設定されたＩＤとが対応付けられた辞書情報を記憶する記憶部と、

前記システムから出力されたテキストログに含まれるメッセージを取得すると、前記記憶部に記憶された辞書情報を参照し、前記テキストログに含まれるメッセージを種別ごとに分類し、分類した各メッセージにＩＤを付与する分類部と、

前記分類部によってメッセージに付与されたＩＤに基づいて、異常を検知する検知部と

を有することを特徴とする異常検知装置。

[請求項2] 前記検知部は、前記分類部によって分類されたメッセージのうち、前記記憶部の辞書情報の中に該当するメッセージの種別が存在しなかったメッセージの種別の件数が所定の閾値を超えている場合には、異常を検知することを特徴とする請求項１に記載の異常検知装置。

[請求項3] 前記検知部は、前記分類部によって前記メッセージが分類された結果から、前記メッセージに付与されたＩＤの単位時間あたりの頻度をＩＤごとに集計し、任意の単位時間の頻度と他の単位時間の頻度とを比較し、各単位時間のＩＤごとの異常度を検知することを特徴とする請求項１に記載の異常検知装置。

[請求項4] 前記分類部は、前記メッセージを所定の区切り文字で分割して単語群を抽出し、各単語を文字種によってパラメータまたは非パラメータに分類し、前記記憶部に記憶された辞書情報の種別群の中で、非パラメータと分類された箇所の単語が一致する種別が存在する場合には、該種別に対応するＩＤを前記メッセージに付与し、非パラメータと分類された箇所の単語が一致する種別が存在しない場合には、前記メッセージに新規ＩＤを付与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異常検知装置。

[請求項5] 前記分類部によって付与されたＩＤを含むメッセージに関する情報を通知する通知部をさらに有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異常検知装置。

[請求項6] 異常検知装置によって実行される異常検知方法であって、

前記異常検知装置は、システムから出力されたテキストログに含まれる各メッセージの種別を表す各メッセージの部分文字列と、メッセージの種別ごとに設定されたＩＤとが対応付けられた辞書情報を記憶する記憶部を有し、

前記システムから出力されたテキストログに含まれるメッセージを取得すると、前記記憶部に記憶された辞書情報を参照し、前記テキストログに含まれるメッセージを種別ごとに分類し、分類した各メッセージにＩＤを付与する分類工程と、

前記分類工程によってメッセージに付与されたＩＤに基づいて、異常を検知する検知工程と

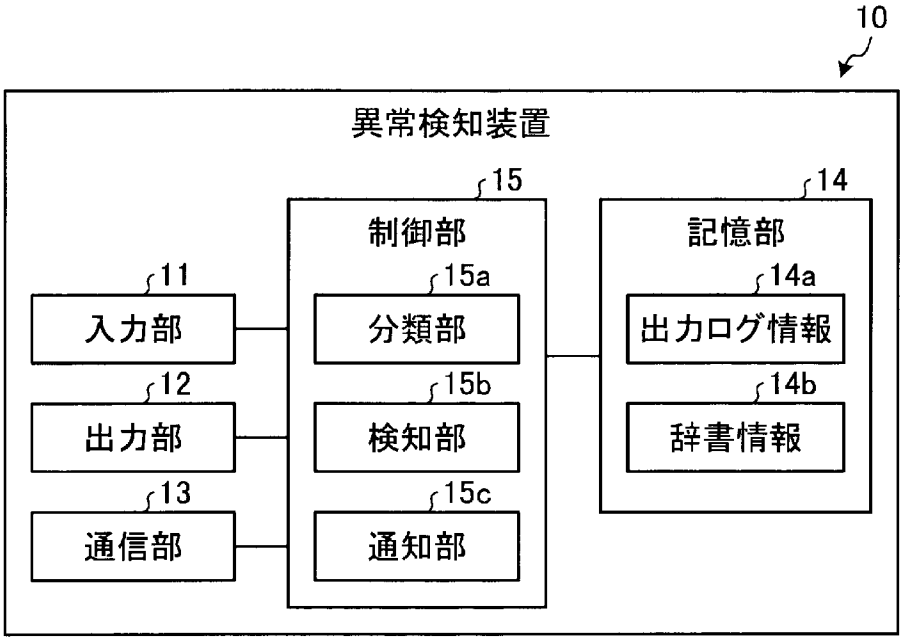
を含むことを特徴とする異常検知方法。

[請求項7] システムから出力されたテキストログに含まれるメッセージを取得すると、前記システムから出力されたテキストログに含まれる各メッセージの種別を表す各メッセージの部分文字列とメッセージの種別ごとに設定されたＩＤとが対応付けられた辞書情報を記憶する記憶部に記憶された前記辞書情報を参照し、前記テキストログに含まれるメッセージを種別ごとに分類し、分類した各メッセージにＩＤを付与する分類ステップと、

前記分類ステップによってメッセージに付与されたＩＤに基づいて、異常を検知する検知ステップと

をコンピュータに実行させることを特徴とする異常検知プログラム。

[図1]



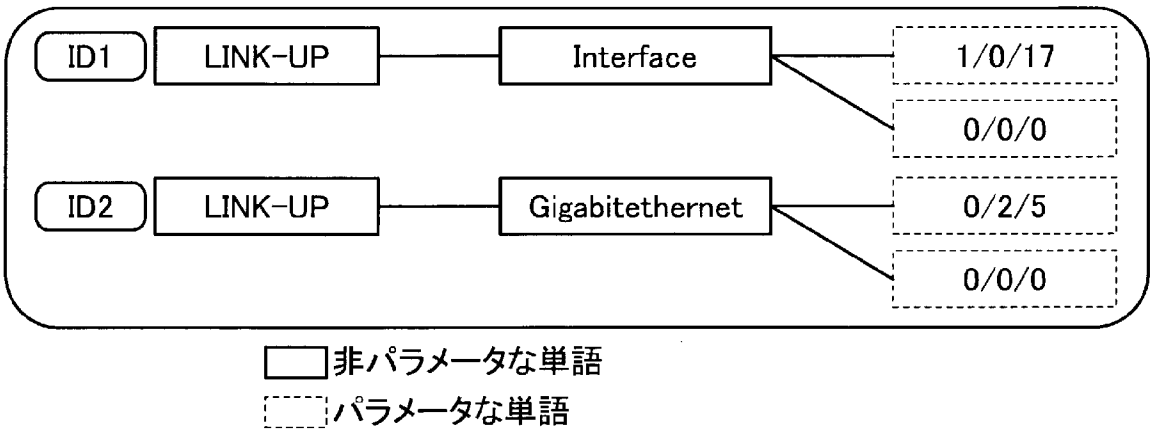
[図2]

2015/05/18T14:56 LINK-UP Interface 1/0/17
2015/05/18T14:58 LINK-UP Interface 0/0/0
2015/05/18T14:59 LINK-UP Gigabitethernet 0/2/5
.....
2015/05/18T15:01 LINK-UP Gigabitethernet 0/0/0

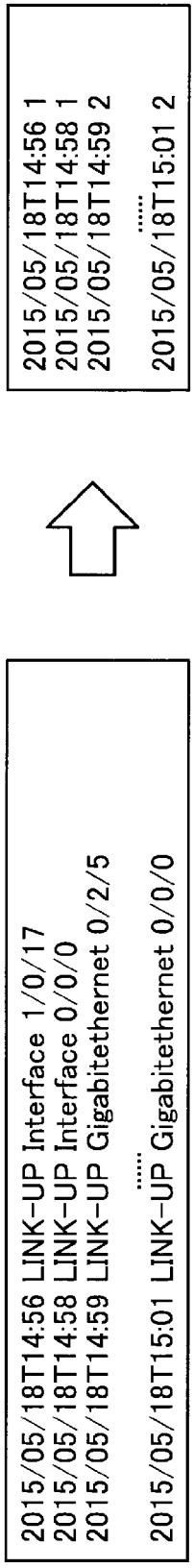
[図3]

ID	単語シーケンス
1	LINK-UP Interface *
2	LINK-UP Gigabitethernet *

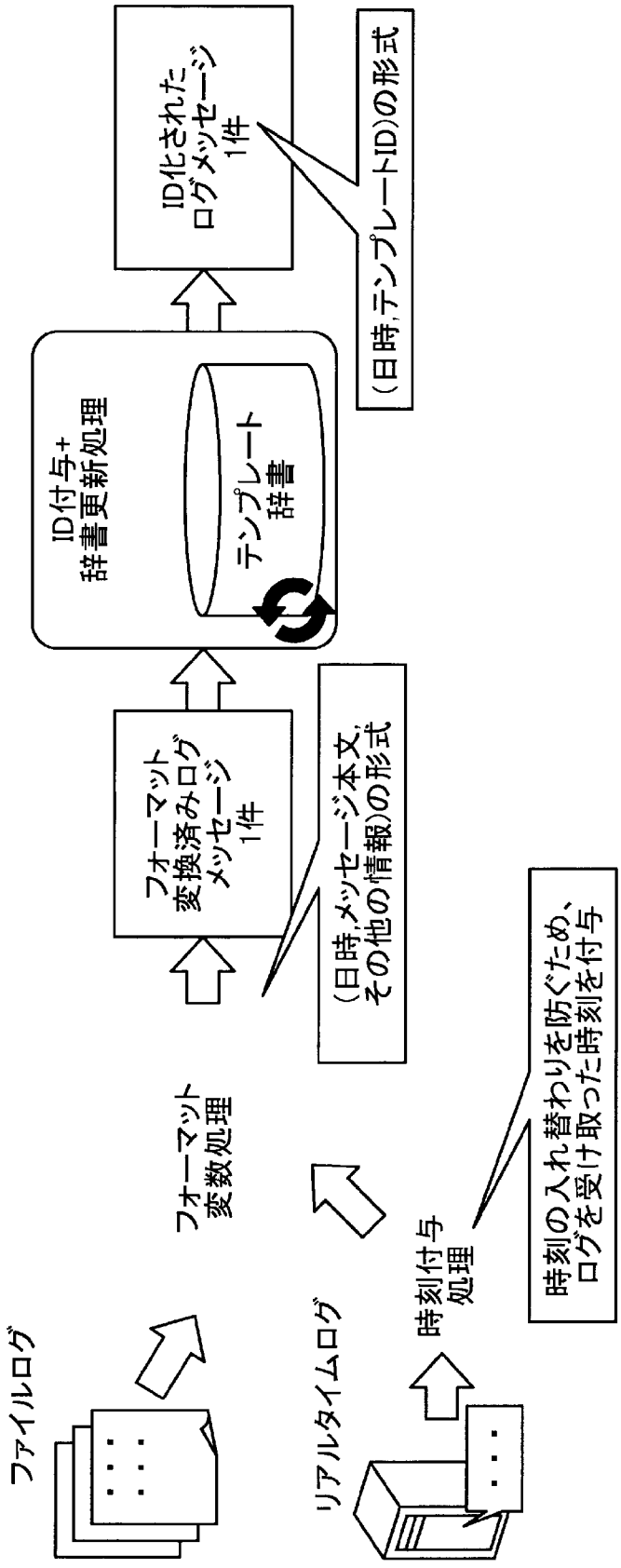
[図4]



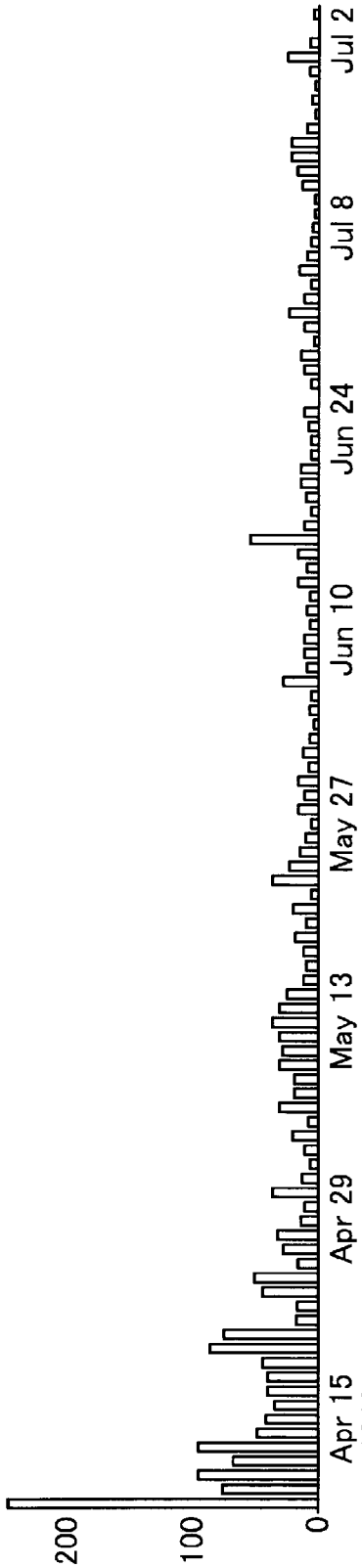
[図5]


2015/05/18T14:56 LINK-UP Interface 1/0/17
2015/05/18T14:58 LINK-UP Interface 0/0/0
2015/05/18T14:59 LINK-UP Gigabitethernet 0/2/5
.....
2015/05/18T15:01 LINK-UP Gigabitethernet 0/0/0
2015/05/18T14:56 1
2015/05/18T14:58 1
2015/05/18T14:59 2
.....
2015/05/18T15:01 2

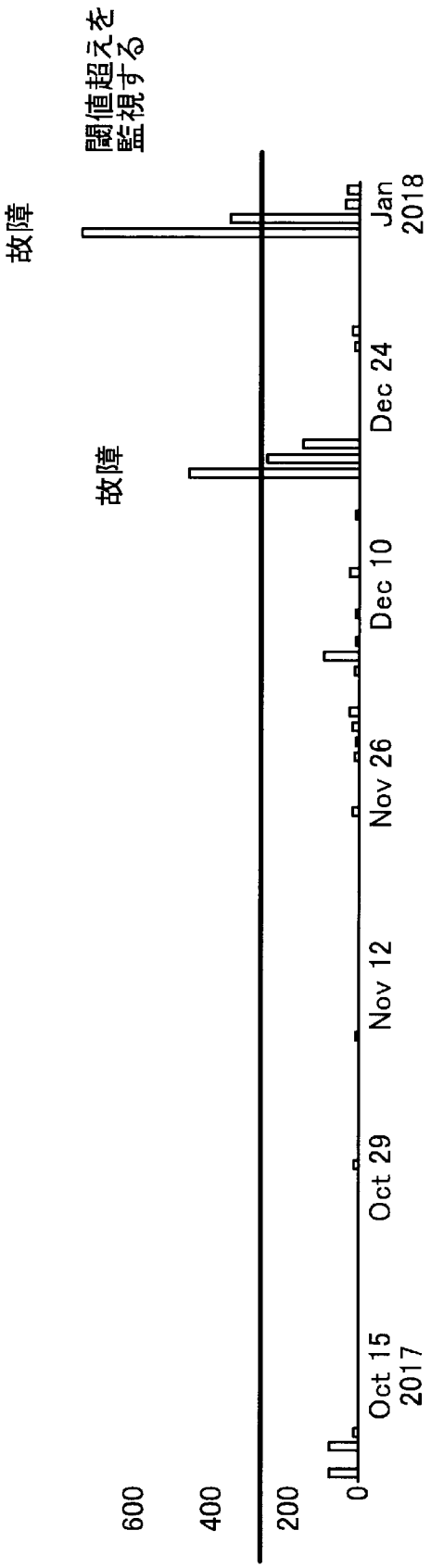
[図6]


ファイルログ
フォーマット
変数処理
フォーマット
変換済みログ
メッセージ
1件
(日時,メッセージ本文,
その他の情報)の形式
ID付与+
辞書更新処理
テンプレート
辞書
(日時,テンプレートID)の形式
ID化された
ログメッセージ
1件
リアルタイムログ
時刻付与
処理
時刻の入れ替わりを防ぐため、
ログを受け取った時刻を付与

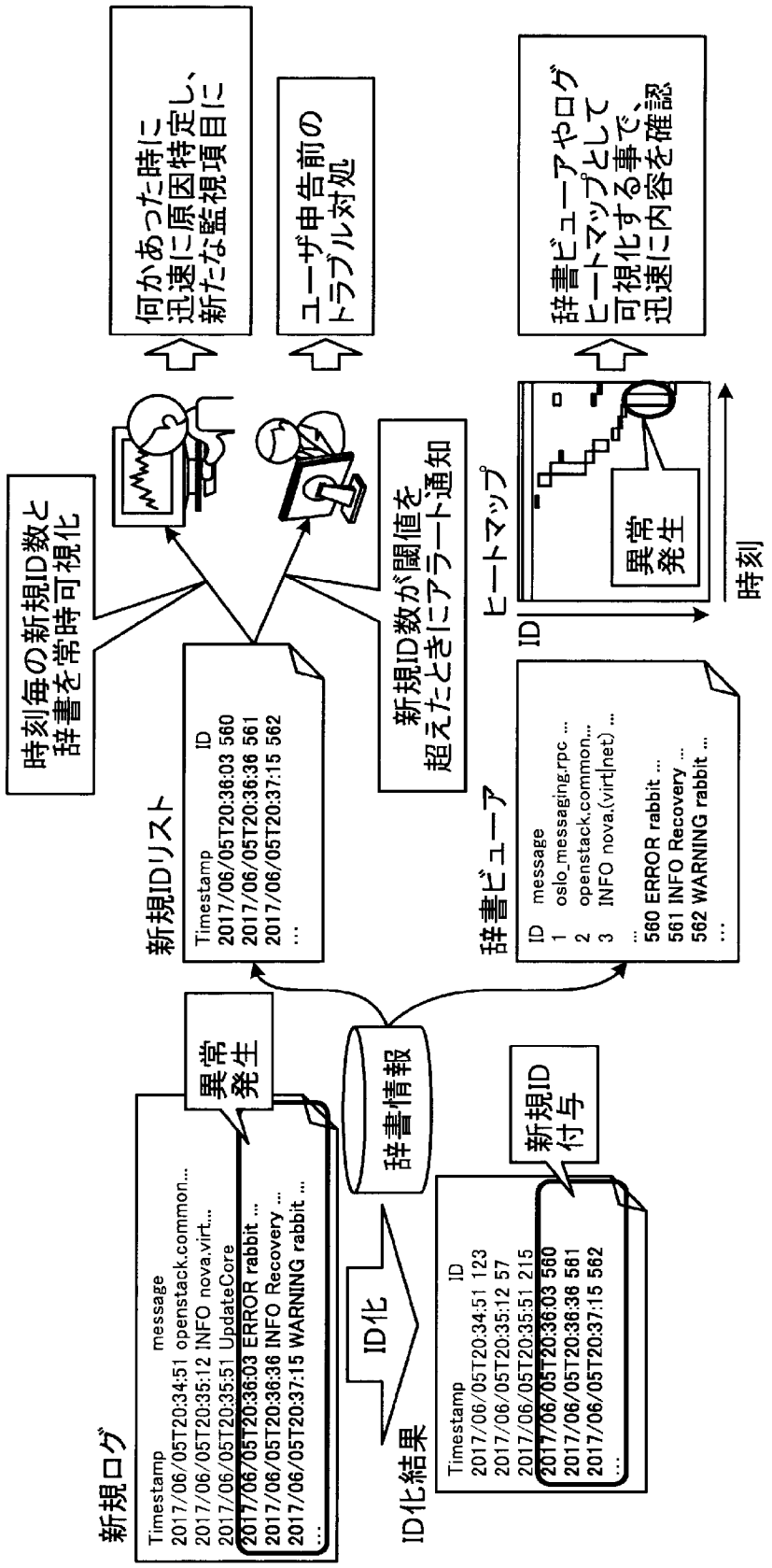
[図7]



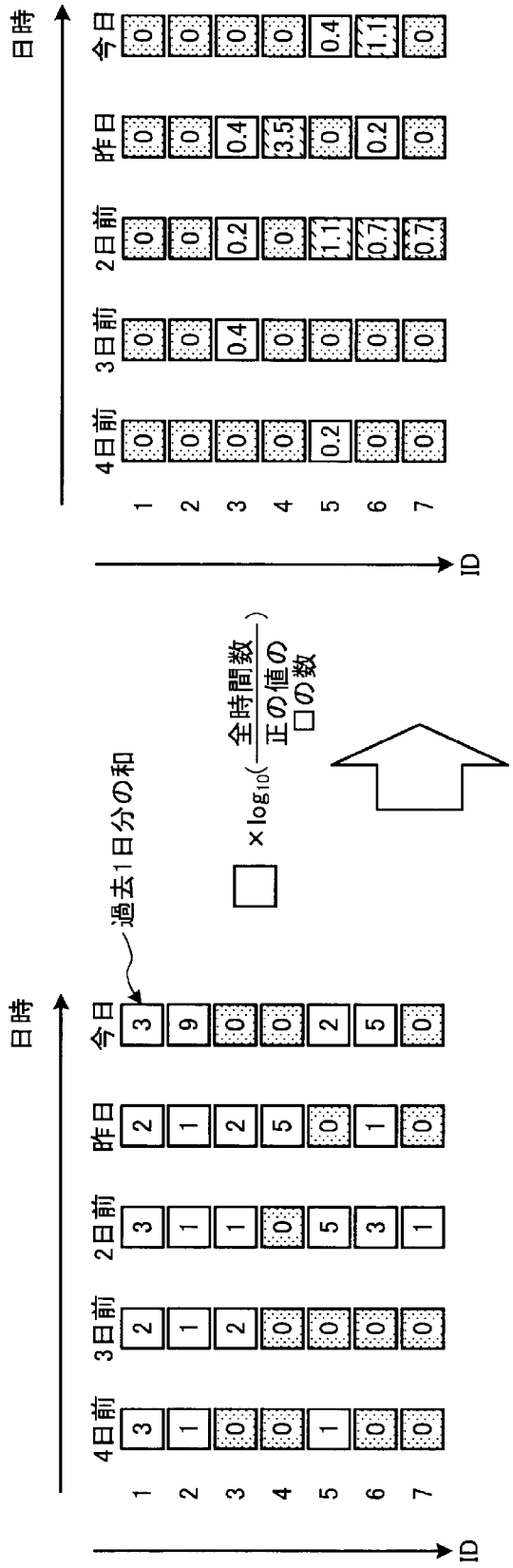
[図8]


閾値超えを監視する
故障
故障
600
400
200
0
Oct 15
2017
Oct 29
Nov 12
Nov 26
Dec 10
Dec 24
Jan
2018

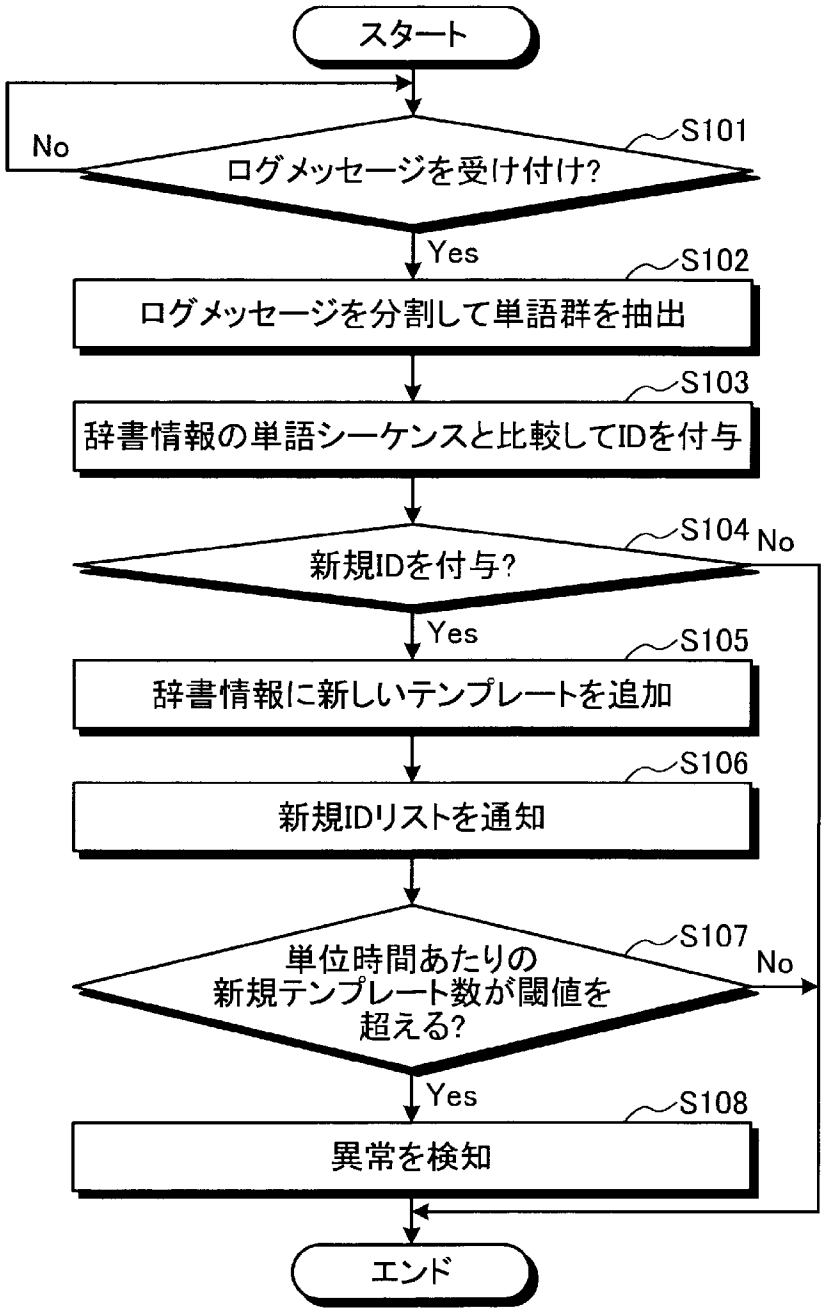
[図9]


新規ログ
Timestamp message
2017/06/05T20:34:51 openstack.common...
2017/06/05T20:35:12 INFO nova.virt...
2017/06/05T20:35:51 UpdateCore
2017/06/05T20:36:03 ERROR rabbit ...
2017/06/05T20:36:36 INFO Recovery ...
2017/06/05T20:37:15 WARNING rabbit ...
...
異常発生
ID化
ID化結果
Timestamp ID
2017/06/05T20:34:51 123
2017/06/05T20:35:12 57
2017/06/05T20:35:51 215
2017/06/05T20:36:03 560
2017/06/05T20:38:36 561
2017/06/05T20:37:15 562
...
新規ID付与
辞書情報
新規IDリスト
Timestamp ID
2017/06/05T20:36:03 560
2017/06/05T20:36:36 561
2017/06/05T20:37:15 562
...
辞書ビューア
ID message
1 oslo_messaging.rpc ...
2 openstack.common...
3 INFO nova.(virt|net) ...
...
560 ERROR rabbit ...
561 INFO Recovery ...
562 WARNING rabbit ...
...
ヒートマップ
ID
異常発生
時刻
時刻毎の新規ID数と辞書を常時可視化
新規ID数が閾値を超えたときにアラート通知
何かあった時に迅速に原因特定し、新たな監視項目に
ユーザ申告前のトラブル対処
辞書ビューアやログヒートマップとして可視化する事で、迅速に内容を確認

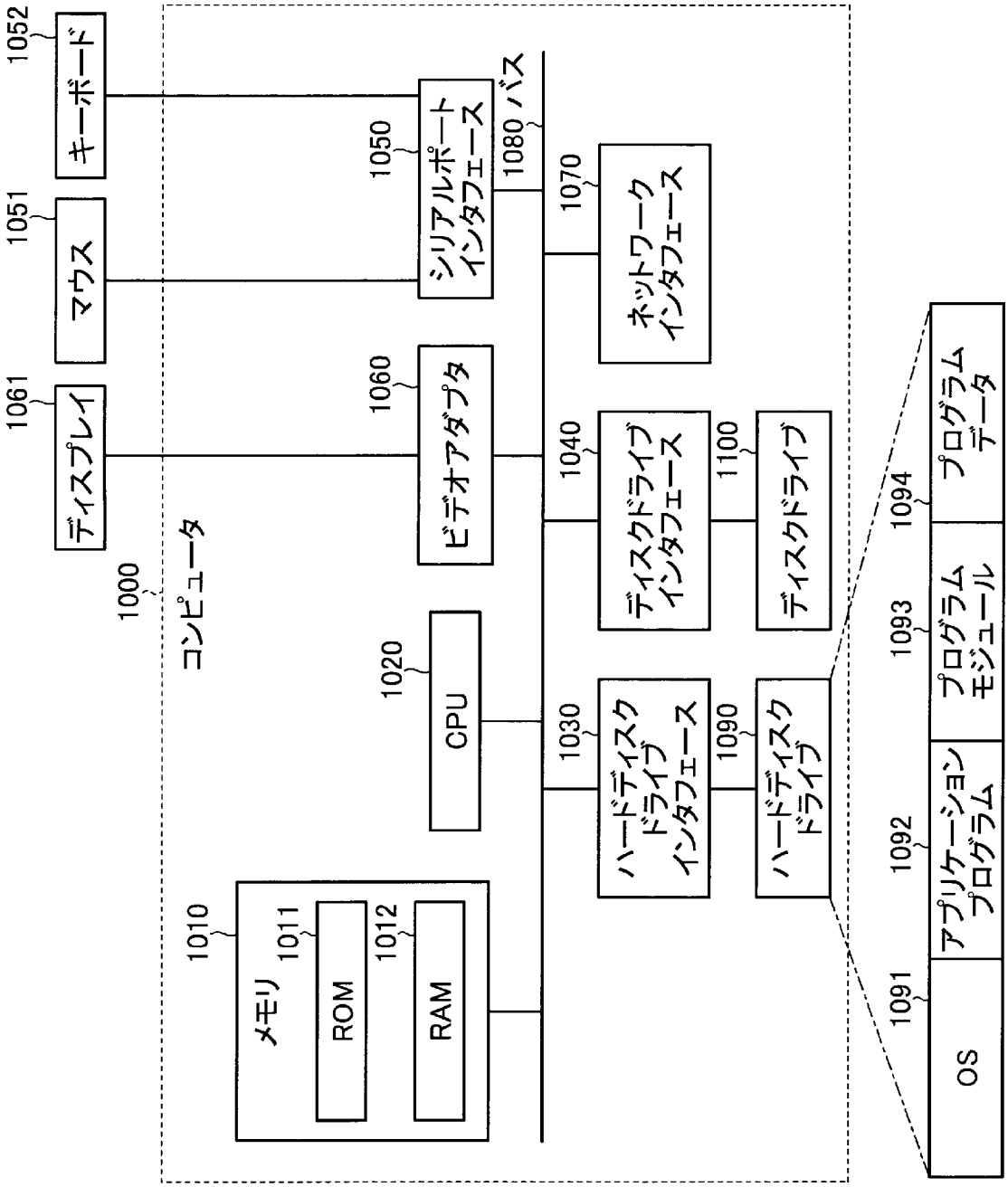
[図10]



[図11]


スタート
S101
No
ログメッセージを受け付け?
Yes
S102
ログメッセージを分割して単語群を抽出
S103
辞書情報の単語シーケンスと比較してIDを付与
S104
No
新規IDを付与?
Yes
S105
辞書情報に新しいテンプレートを追加
S106
新規IDリストを通知
S107
No
単位時間あたりの
新規テンプレート数が閾値を
超える?
Yes
S108
異常を検知
エンド

[図12]


1000
コンピュータ
1010
メモリ
1011
ROM
1012
RAM
1020
CPU
1030
ハードディスクドライブインタフェース
1040
ディスクドライブインタフェース
1050
シリアルポートインタフェース
1060
ビデオアダプタ
1070
ネットワークインタフェース
1080
バス
1090
ハードディスクドライブ
1100
ディスクドライブ
1051
マウス
1052
キーボード
1061
ディスプレイ
1091
OS
1092
アプリケーションプログラム
1093
プログラムモジュール
1094
プログラムデータ

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2020/004969

A. CLASSIFICATION OF SUBJECT MATTER
G06F 11/07(2006.01)i
FI: G06F11/07 160; G06F11/07 140A

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)
G06F11/07

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	1971-2020
Registered utility model specifications of Japan	1996-2020
Published registered utility model applications of Japan	1994-2020

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	WO 2017/081865 A1 (NEC CORP.) 18.05.2017 (2017-05-18) paragraphs [0074]-[0082], [0148], fig. 5-7	1,4,6-7 2-3,5
Y	高田 哲司 他，見えログ：情報視覚化とテキストマイニングを用いたログ情報ブラウザ，情報処理学会論文誌 IPSJ Journal, 10 December 2000, vol. 41, no. 12, pp. 3265-3275, ISSN:0387-5806, page 3266, right column, lines 1-14, (TAKADA, Tetsuji et al., "MieLog : Log Information Browse with Information Visualization and Text Mining")	2-3,5
A	JP 2004-318552 A (KDDI CORP.) 11.11.2004 (2004-11-11) paragraphs [0060]-[0066], fig. 6	1-7

☐ 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	Date of mailing of the international search report
30 March 2020 (31.03.2020)	07 April 2020 (07.04.2020)

Name and mailing address of the ISA/	Authorized officer
Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2020/004969

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
WO 2017/081865 A1	18 May 2017	US 2018/0357214 A1 paragraphs [0095]-[0103], [0169], fig. 5-7	
JP 2004-318552 A	11 Nov. 2004	US 2004/0250169 A1 paragraphs [0091]-[0099], fig. 6	

国際調査報告

国際出願番号
PCT/JP2020/004969

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

G06F 11/07(2006.01)i
FI: G06F11/07 160; G06F11/07 140A

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））

G06F11/07

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	１９７１－２０２０年
日本国実用新案登録公報	１９９６－２０２０年
日本国登録実用新案公報	１９９４－２０２０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	WO 2017/081865 A1 (日本電気株式会社) 18.05.2017 (2017-05-18) 段落0074-0082,0148,図5-7	1,4,6-7
Y		2-3,5
Y	高田 哲司 他, 見えログ：情報視覚化とテキストマイニングを用いたログ情報ブラウザ, 情報処理学会論文誌 IPSJ Journal, 2000.12.10, 第41巻 第12号, 第3265頁-第3275頁, ISSN:0387-5806 第3266頁右欄第1行目-第14行目	2-3,5
A	JP 2004-318552 A (ＫＤＤＩ株式会社) 11.11.2004 (2004-11-11) 段落0060-0066,図6	1-7

☐ 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抵触するものではなく、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	国際調査報告の発送日
30.03.2020	07.04.2020

名称及びあて先

日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目４番３号

権限のある職員（特許庁審査官）

三坂　敏夫 5B 4178

電話番号 03-3581-1101 内線 3545

国際調査報告
パテントファミリーに関する情報

国際出願番号
PCT/JP2020/004969

引用文献	公表日	パテントファミリー文献	公表日
WO 2017/081865 A1	18.05.2017	US 2018/0357214 A1 Paragraphs0095-0103,0169,Figures5-7	
JP 2004-318552 A	11.11.2004	US 2004/0250169 A1 Paragraphs0091-0099,Figure6	