

(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
2015年4月2日(02.04.2015)



(10) 国際公開番号
WO 2015/045262 A1

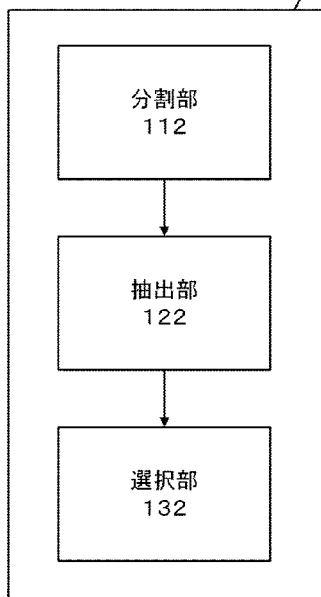
- (51) 国際特許分類:
G06F 11/34 (2006.01)
- (21) 国際出願番号: PCT/JP2014/004275
- (22) 国際出願日: 2014年8月21日(21.08.2014)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2013-197416 2013年9月24日(24.09.2013) JP
- (71) 出願人: 日本電気株式会社(NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).
- (72) 発明者: 外川 遼介(TOGAWA, Ryosuke); 〒1088001 東京都港区芝五丁目7番1号日本電気株式会社内 Tokyo (JP).
- (74) 代理人: 下坂 直樹(SHIMOSAKA, Naoki); 〒1088001 東京都港区芝五丁目7番1号日本電気株式会社内 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, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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), ユーロパ (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: LOG ANALYSIS SYSTEM, FAULT CAUSE ANALYSIS SYSTEM, LOG ANALYSIS METHOD, AND RECORDING MEDIUM WHICH STORES PROGRAM

(54) 発明の名称: ログ分析システム、障害原因分析システム、ログ分析方法、および、プログラムを記憶する記録媒体

ログ分析システム102



102 Log analysis system
112 Segmenting unit
122 Extraction unit
132 Selection unit

(57) Abstract: An objective of the present invention is to appropriately extract, or assist in extracting, a normal operation pattern from a normal operation log. A log file including information wherein a plurality of log records is ordered in time series is segmented into a plurality of log groups according to prescribed rules. A plurality of patterns which are configured by a plurality of events which are contiguous in time series is extracted from the plurality of log groups, and the patterns are associated with frequency information which represents how many log records the pattern is extracted from. The pattern to be outputted from the plurality of patterns is selected on the basis of an inclusion relation between the plurality of events which are contiguous in time series and configure a given pattern and the plurality of events which are contiguous in time series and configure other patterns, and on the basis of the frequency information which is associated with the patterns.

(57) 要約: 正常ログから正常パターンを適切に抽出すること、または抽出することを支援する。複数のログレコードが時系列順に整列された情報を含むログファイルを、所定のルールに従って、複数のロググループへと分割し、複数の前記ロググループから、時系列的に連続する複数のイベントによって構成される複数のパターンを抽出し、前記パターンを、当該パターンが幾つのログレコードから抽出されたかを表す頻度情報と関連づけ、あるパターンを構成する時系列的に連続する複数のイベントと、他のパターンを構成する時系列的に連続した複数のイベントとの包含関係に基づいて、かつ、前記パターンに関連づけられた頻度情報に基づいて、前記複数のパターンから出力すべきパターンを選択する。

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：

ログ分析システム、障害原因分析システム、ログ分析方法、および、プログラムを記憶する記録媒体

技術分野

[0001] 本発明は、情報処理システムが出力するログを分析する技術に関する。

背景技術

[0002] 情報処理システムを運用保守する際には、運用保守を行う者（以下、「管理者」と記載する）は、当該情報処理システムに障害が発生した場合、当該障害の原因を特定する必要がある。管理者が障害の原因を特定する際に、当該情報処理システムが出力したログは重要な手がかりとなり得る。しかし、情報処理システムが出力するログのサイズは膨大であるので、管理者はその全てを詳細に分析することはできない。

[0003] そのため、管理者が障害の原因を特定する際には、情報処理システムが出力する膨大な量のログから、当該障害が発生した直接的または間接的な原因を示すメッセージ（以下、「原因メッセージ」と記載する）を特定する必要がある。そして管理者は、特定された原因メッセージ、および、当該原因メッセージの周辺のログを詳細に分析することにより、障害の原因を特定する。

[0004] 特許文献１は、情報処理システムが出力する膨大なサイズのログから、正常動作時とは異なる頻度で出現するメッセージのパターンを検出する技術の一例を開示する。特許文献１が開示する技術は、情報処理システムが正常に動作していた際に当該情報処理システムが出力したログ（以下、「正常ログ」と記載する）から、正常ログにおけるメッセージの出現頻度等の典型的なパターン（以下、「正常パターン」と記載する）を抽出する。特許文献１が開示する技術は、正常パターンと、当該情報処理システムが出力した分析対象のログ（以下、「分析対象ログ」）とを比較し、両者の差分を抽出する。

これにより特許文献１が開示する技術は、分析対象ログから正常動作時とは異なる頻度で出現するメッセージを検出することができる。

先行技術文献

特許文献

[0005] 特許文献１：特許第４９４４３９１号公報

特許文献２：特開２０１２－１９４６０４号公報

特許文献３：特開２０１２－１６８７０２号公報

発明の概要

発明が解決しようとする課題

[0006] 特許文献１は、正常ログから正常パターンをどのように抽出するかについて、詳細には開示していない。情報処理システムに発生した障害の原因分析が適切に行われるためには、正常ログから正常パターンが適切に抽出されることが必要である。

[0007] 「正常ログから正常パターンを適切に抽出する」ということについて説明する。「正常ログから正常パターンを適切に抽出する」ということは、正常ログに基づいて、情報処理装置においてどのような処理がどれくらいの頻度で実行されたかを示す情報を抽出する、ということである。

[0008] 情報処理装置において何らかの処理が実行されると、情報処理装置は、当該処理が実行されたことに応じて複数のメッセージを所定の時系列順に出力する。以降、情報処理装置が特定の処理を実行したことに応じて当該情報処理装置が出力した所定の時系列順の複数のメッセージを、「一連のメッセージ」と記載する。

[0009] 正常ログには、膨大な数量のメッセージが含まれる。正常ログに基づいて、どのような処理がどれくらいの頻度で実行されたかを示す情報を抽出するためには、「一連のメッセージを処理単位でグループ化する情報」が必要である。

[0010] 特許文献２および特許文献３は、「一連のメッセージを処理単位でグルー

プ化する情報」に基づいて、ログを分析する技術を開示する。

[0011] 特許文献2が開示する技術は、メッセージに含まれるトランザクションIDを「一連のメッセージを処理単位でグループ化する情報」として用いる。特許文献2が開示する技術は、メッセージに含まれるトランザクションIDに基づいて、同一のトランザクションIDを含む複数のメッセージを、当該トランザクションに関連する一連のメッセージとしてグルーピングする。

[0012] 特許文献3が開示する技術は、プロセスに関連する一連のメッセージを事前に定義した情報を、「一連のメッセージを処理単位でグループ化する情報」として用いる。特許文献3が開示する技術は、特許文献3が開示する技術は、当該情報に基づいて、複数のメッセージを、特定のプロセスに関連する一連のメッセージとしてグルーピングする。

[0013] しかし、正常ログから正常パターンを適切に抽出する場合において、「一連のメッセージを処理単位でグループ化する情報」は、必ずしも事前に与えられるとは限らない。特許文献2が開示する技術において、メッセージにトランザクションIDが付与されていない状況も考えられる。

[0014] 本発明は、一連のメッセージを処理単位でグループ化する情報が事前に与えられることなしに、正常ログから正常パターンを適切に抽出すること、または抽出することを支援することを目的の1つとする。

課題を解決するための手段

[0015] 本発明の第1の側面は、イベントと当該イベントが発生した時刻とが関連づけられた情報である複数のログレコードが時系列順に整列された情報を含むログファイルを、所定のルールに従って、時系列順に連続する複数のログレコードの集合である複数のロググループへと分割する分割手段と、複数の前記ロググループから、時系列的に連続する複数のイベントによって構成される複数のパターンを抽出し、前記パターンを、当該パターンが幾つのログレコードから抽出されたかを表す頻度情報と関連づける抽出手段と、あるパターンを構成する時系列的に連続する複数のイベントと、他のパターンを構成する時系列的に連続した複数のイベントとの包含関係に基づいて、かつ、

前記パターンに関連づけられた頻度情報に基づいて、前記複数のパターンから出力すべきパターンを選択する選択手段と、を備えるログ分析システムである。

[0016] 本発明の第2の側面は、コンピュータが、イベントと当該イベントが発生した時刻とが関連づけられた情報である複数のログレコードが時系列順に整理された情報を含むログファイルを、所定のルールに従って、時系列順に連続する複数のログレコードの集合である複数のロググループへと分割し、複数の前記ロググループから、時系列的に連続する複数のイベントによって構成される複数のパターンを抽出し、前記パターンを、当該パターンが幾つのログレコードから抽出されたかを表す頻度情報と関連づけ、あるパターンを構成する時系列的に連続する複数のイベントと、他のパターンを構成する時系列的に連続した複数のイベントとの包含関係に基づいて、かつ、前記パターンに関連づけられた頻度情報に基づいて、前記複数のパターンから出力すべきパターンを選択する、ログ分析方法である。

[0017] 本発明の第3の側面は、コンピュータに、イベントと当該イベントが発生した時刻とが関連づけられた情報である複数のログレコードが時系列順に整理された情報を含むログファイルを、所定のルールに従って、時系列順に連続する複数のログレコードの集合である複数のロググループへと分割する処理と、複数の前記ロググループから、時系列的に連続する複数のイベントによって構成される複数のパターンを抽出し、前記パターンを、当該パターンが幾つのログレコードから抽出されたかを表す頻度情報と関連づける処理と、あるパターンを構成する時系列的に連続する複数のイベントと、他のパターンを構成する時系列的に連続した複数のイベントとの包含関係に基づいて、かつ、前記パターンに関連づけられた頻度情報に基づいて、前記複数のパターンから出力すべきパターンを選択する処理と、を実行させるプログラムを記録する記憶媒体である。

[0018] また、本発明の目的は、上記のプログラムが格納されたコンピュータ読み取り可能な記録媒体に加えて、その記憶媒体に格納されているプログラムに

よっても達成される。

発明の効果

[0019] 本発明によれば、一連のメッセージを処理単位でグループ化する情報が事前に与えられることなしに、正常ログから正常パターンを適切に抽出すること、または抽出することを支援することができる。

図面の簡単な説明

[0020] [図1]図1は、本発明の第1の実施形態にかかるログ分析システム100を説明するブロック図である。

[図2]図2は、本発明の第1の実施形態にかかるログ分析システム100に入力される正常ログ910の一例を表す図である。

[図3]図3は、本発明の第1の実施形態にかかるログ分析システム100の構成を表すブロック図である。

[図4]図4は、本発明の第1の実施形態にかかる分割部110によって分割されたログファイルの一例を表す図である。

[図5]図5は、本発明の第1の実施形態にかかる抽出部120によって抽出されたパターンの一例を表す図である。

[図6]図6は、本発明の第1の実施形態にかかる選択部130の動作を説明する図である。

[図7]図7は、本発明の第1の実施形態にかかる選択部130の動作を説明する図である。

[図8]図8は、本発明の第1の実施形態にかかるログ分析システム100の動作を説明するフローチャートである。

[図9]図9は、ログ分析システム100を実現するハードウェア構成の一例を表す図である

[図10]図10は、本発明の第2の実施形態にかかるログ分析システム101の構成を表すブロック図である。

[図11A]図11Aは、本発明の第2の実施形態にかかる変換部150に入力されるログファイルの一例を表す第1の図である。

[図11B]図 1 1 B は、本発明の第 2 の実施形態にかかる変換部 1 5 0 に入力されるログファイルの一例を表す第 2 の図である。

[図11C]図 1 1 C は、本発明の第 2 の実施形態にかかる変換部 1 5 0 に入力されるログファイルの一例を表す第 3 の図である。

[図12]図 1 2 は、本発明の第 2 の実施形態にかかるフォーマットの一例を表す図である。

[図13]図 1 3 は、本発明の第 2 の実施形態にかかるフォーマットの一例を表す図である。

[図14]図 1 4 は、本発明の第 2 の実施形態にかかるログ分析システム 1 0 1 の動作を説明するフローチャートである。

[図15]図 1 5 は、本発明の第 3 の実施形態にかかる障害原因分析システム 3 0 0 の構成を表すブロック図である。

[図16]図 1 6 は、本発明の第 4 の実施形態にかかる第 4 の実施形態にかかるログ分析システム 1 0 2 の動作を説明するフローチャートである。

発明を実施するための形態

[0021] 以下、本発明の実施形態について、図面を参照して詳細に説明する。

[0022] <第 1 の実施形態>

図 1 は、本発明の第 1 の実施形態にかかるログ分析システム 1 0 0 を説明するブロック図である。図 1 に示すように、ログ分析システム 1 0 0 は、情報処理システム 9 0 0 から出力された正常ログ 9 1 0 を取得する。上述したように、正常ログ 9 1 0 とは、情報処理システム 9 0 0 が正常に動作していた際に当該情報処理システム 9 0 0 が出力したログである。ログ分析システム 1 0 0 は、正常ログ 9 1 0 に基づいて、情報処理システム 9 0 0 においてどのような処理がどれくらいの頻度で実行されたかを示す情報、すなわち正常パターン 9 2 0 を抽出する。なお、後述するように、正常パターン 9 2 0 においては、必ずしも具体的な処理が特定される必要はない。ログ分析システム 1 0 0 は、正常パターン 9 2 0 を、正常パターン記憶部 2 0 0 に出力する。以下の説明において、「情報処理システム 9 0 0」は、本実施形態に係

るログ分析システム１００及び本発明の他の実施形態に係るログ分析システムによって取得される、正常ログ９１０を出力する情報処理装置である。「情報処理システム９００」本発明の他の実施形態に係る障害原因分析システム３００によって取得される、分析対象ログ９９０を出力する情報処理装置でもある。

[0023] 図２は、情報処理システム９００から出力された正常ログ９１０の一例を説明する図である。正常ログ９１０は、ログファイルとも呼ばれる。正常ログ９１０は、複数のログレコードが時系列順に整列された情報を含む。

[0024] ログレコードは、情報処理システム９００において発生したイベントと当該イベントが発生した時刻とが関連づけられた情報である。ログレコードは、メッセージとも呼ばれる。ログレコードは、図２に示した「行」に相当する情報である。図２の１行目に示したログレコードを参照すると、情報処理システム９００において、イベントＩＤ（識別子）「Ａ」によって識別されるイベントが時刻０９：０４：０１に発生したことがわかる。図２の２行目に示したログレコードを参照すると、情報処理システム９００において、イベントＩＤ（識別子）「Ｂ」によって識別されるイベントが時刻０９：０４：０２に発生したことがわかる。イベントＩＤは、情報処理システム９００において発生したイベントを識別する識別子である。

[0025] 以降、実施形態を説明する際に、イベントを特定する必要がある場合には、イベントＩＤを用いて特定する。また、以降、例えば「ある時刻においてイベントＡが発生したことを示すログレコード」を、「イベントＡのログレコード」と表現する場合がある。

[0026] 情報処理システム９００が何らかの処理を実行すると、情報処理システム９００は、当該処理が実行されたことに応じて複数のイベントを所定の時系列順に発生する。以降、情報処理システム９００が特定の処理を実行したことに応じて当該情報処理システム９００が発生する所定の時系列順の複数のイベントを、「一連のイベント」と記載する。

[0027] 理解を容易にするため、以下に示すような具体的な前提において、本実施

形態を説明する。情報処理システム 900 が（処理 1）を実行した場合、情報処理システム 900 は、イベント A、イベント B および イベント C をこの順番で発生することとする。情報処理システム 900 が（処理 2）を実行した場合、情報処理システム 900 は、イベント X、および イベント Y をこの順番で発生することとする。

[0028] 従って、情報処理システム 900 は、（処理 1）を実行した場合、イベント A のログレコード、イベント B のログレコードおよび イベント C のログレコードを、この順番で出力する。情報処理システム 900 は、（処理 2）を実行した場合、イベント X のログレコードおよび イベント Y のログレコードを、この順番で出力する。

[0029] 正常ログ 910 には、膨大な数量のログレコードが含まれる。ログ分析システム 100 は、正常ログ 910 に基づいてどのような処理がどれくらいの頻度で実行されたかを示す情報を抽出する。例えば、ログ分析システム 100 は、「所定時間の間に処理 1 が 10 回実行され処理 2 が 18 回実行された」というような情報を正常パターン 920 として抽出する。なお、正常パターン 920 にあっては、必ずしも具体的な処理が特定される必要はない。例えば、正常パターン 920 は「所定時間の間にイベント A→B→C というパターンが 10 回発生し、イベント X→Y というパターンが 18 回発生した」というような情報でもよい。正常パターン 920 がこのような情報であっても、情報処理システム 900 に発生した障害の原因を分析する際の比較対象としての役割は十分に果たすことができる。

[0030] なお、記号「→」は、当該記号の前後に発生したイベントが、正常ログ 910 において時系列的に連続していることを示す記号である。すなわち、「A→B→C」とは、正常ログ 910 においてイベント A のログレコードとイベント B のログレコードとイベント C のログレコードとがこの順番で連続していることを示す。以降、複数のイベントが連続して発生したことを表す際に、記号「→」を省略して、例えば「A→B→C」を「ABC」と表現する場合がある。

[0031] ログ分析システム１００が正常ログ９１０に基づいてこのような情報を抽出するためには、「一連のログレコードを処理単位でグループ化する情報」が必要である。

[0032] ここで、「一連のログレコードを処理単位でグループ化する情報」は、例えば、「イベントＡのログレコード、イベントＢのログレコードおよびイベントＣのログレコードが何らかの処理に関連付いている」あるいは「イベントＸのログレコードおよびイベントＹのログレコードが何らかの処理に関連付いている」という情報である。

[0033] 本実施形態にかかるログ分析システム１００は、事前に「一連のログレコードを処理単位でグループ化する情報」を明示的に与えられることなく、正常ログ９１０に基づいて正常パターン９２０を適切に抽出できる。なぜなら、本実施形態にかかるログ分析システム１００は、正常ログ９１０に基づいて、「一連のログレコードを処理単位でグループ化する情報」を自ら抽出することができるからである。

[0034] 言い換えると、ログ分析システム１００は、事前に「イベントＡのログレコード、イベントＢのログレコードおよびイベントＣのログレコードが何らかの処理に関連付いている」あるいは「イベントＸのログレコードおよびイベントＹのログレコードが何らかの処理に関連付いている」という情報を与えられていなくても、正常ログ９１０に基づいて、このような情報を自ら抽出することができる、ということである。

[0035] 本実施形態にかかる情報処理システム９００は、前提として、下記に示すような性質を有するものとする。

[0036] （前提１）情報処理システム９００は、複数の処理を並行して実行する。

[0037] （前提２）例えば情報処理システム９００が、処理１と処理２とを並列に実行すると、情報処理システム９００が出力するログファイルにおいて処理１に関連するイベントのログレコードと、処理２に関連するログレコードとが、時系列的に混ざって出力される場合がある。例えば、図２に示す０９：０４：１２から０９：０４：１６までに出力されたログレコードにおいては

、処理 1 に関連するイベントのログレコードと、処理 2 に関連するイベントのログレコードとが、時系列的に混ざっている。図 2 に示す例では、イベント $A \rightarrow B \rightarrow C$ が、処理 1 に関連するイベントである。イベント $X \rightarrow Y$ が、処理 2 に関連するイベントである。

[0038] (前提 3) 上述したような、ある処理に関連するイベントのログレコードと、他の処理に関連するイベントのログレコードとが、時系列的に混ざって出力されることは、稀である。ほとんどの場合、特定の処理に関連するイベントのログレコードは、ログファイルにおいて時系列的に連続となる。

[0039] かかる前提に基づいて、ログ分析システム 100 が、正常ログ 910 からどのように正常パターン 920 を適切に抽出するかを、以下、詳細に説明する。

[0040] 図 3 は、図 1 に示すログ分析システム 100 の構成を示すブロック図である。図 3 に示すように、ログ分析システム 100 は、分割部 110 と、抽出部 120 と、選択部 130 と、正常パターン抽出部 140 と、を備える。

[0041] 分割部 110 は、正常ログ 910 を取得し、取得した正常ログ 910 を複数のロググループへと分割する。ロググループは、時系列的に連続する複数のログレコードの集合である（詳細は後述する）。

[0042] 抽出部 120 は、複数のロググループから、時系列的に連続する複数のイベントから構成される複数の「パターン」を抽出する。抽出部 120 は、パターンを、当該パターンが幾つのログレコードから抽出されたかを表す頻度情報と関連づける（詳細は後述する）。

[0043] 選択部 130 は、あるパターンを構成する時系列的に連続する複数のイベントと他のパターンを構成する時系列的に連続した複数のイベントとの包含関係に基づいて、前記複数のパターンから出力するパターンを選択する。また、選択部 130 は、前記パターンに関連づけられた頻度情報に基づいて、前記複数のパターンから出力するパターンを選択する（詳細は後述する）。

[0044] 正常パターン抽出部 140 は、選択部 130 が出力したパターンに基づいて、正常ログ 910 から正常パターン 920 を抽出する。

- [0045] 分割部 110 の詳細を説明する。図 4 は、図 2 に示す正常ログ 910 が、分割部 110 によって複数のロググループに分割された様子を説明する図である。図 4 に示すロググループは、分割部 110 が、5 秒間隔で、正常ログ 910 をロググループに分割した場合における、ロググループである。
- [0046] 分割部 110 は、あらかじめ指定された所定の時間間隔で正常ログ 910 を分割する。好ましくは、分割部 110 は、情報処理システム 900 が実行する処理に応じて出力される一連のイベントのログレコードが、一つのロググループに包含されると見込まれる程度の時間間隔で、正常ログ 910 を分割することが望ましい。情報処理システム 900 が特定の処理を実行するのに要する時間は、情報処理システム 900 の負荷状況等によって毎回異なるが、分割部 110 は例えば、平均処理時間等の値を算出し、このような値に基づいて分割の時間幅を決定することが望ましい。
- [0047] 分割の時間幅を決定する際には、プロセスまたはトランザクションの実行時間のポアソン分布等の値が参酌されてもよい。
- [0048] 抽出部 120 の詳細を説明する。図 5 は、抽出部 120 によって複数のロググループから抽出された「パターン」の一例と、当該パターンが抽出された「頻度」とを表す図である。「パターン」とは、ロググループの中に見られる、時系列的に連続する複数のイベントから構成される組み合わせである。「頻度」とは、当該パターンが幾つのログレコードから抽出されたかを表す情報である。
- [0049] 具体例を用いて説明する。例えば、抽出部 120 が、図 4 に示すロググループ 1 からパターンを抽出する場合を考える。ロググループ 1 を構成する時系列的に連続する複数のログレコードは、すなわち「A B C X Y」である。抽出部 120 が、ロググループ 1 から抽出し得るパターンを、以下に列挙する。「A B C X Y」、「A B C X」、「B C X Y」、「A B C」、「B C X」、「C X Y」、「A B」、「B C」、「C X」、「X Y」、「A」、「B」、「C」、「X」、「Y」。
- [0050] 抽出部 120 は、複数のロググループから「パターン」を抽出する。抽出

部 1 2 0 は、パターンを、当該パターンが幾つのログレコードから抽出されたかを表す頻度情報と関連づける。例えば図 5 において、パターン「A B C」の頻度は 4 3 である。これは、パターン「A B C」は、4 3 個のロググループから抽出されたことを意味する。

[0051] 抽出部 1 2 0 がパターンを抽出する具体的な方法としては、さまざまなアルゴリズムを用いることができる。抽出部 1 2 0 は例えば、apriori、prefix span、LCM (Linear time Closed itemset Miner)、BIDE (Bi-Directional Execution) もしくは FP (Frequent Pattern)-growth、またはこれらのアルゴリズムの応用を用いて、パターンを抽出する。

[0052] 以降、情報処理システム 9 0 0 が実行する特定の処理に関連付くパターンを、「意味のあるパターン」と記載する。例えば、パターン「A B C」は処理 1 に関連するパターンなので、意味のあるパターンである。例えば、パターン「X Y」は処理 2 に関連するパターンなので、意味のあるパターンである。

[0053] 以降、情報処理システム 9 0 0 が実行する特定の処理に関連付かないパターンを、「意味の無いパターン」と記載する。意味の無いパターンの 1 つの類型は、例えば、意味のあるパターンの部分集合に過ぎないパターンである。例えば、「A B」または「B C」などのパターンは、独立してそれ自身何らかの処理に関連付いているわけではないので、意味のないパターンである。意味の無いパターンの他の類型は、例えば、意味のあるパターンに何らかのイベントが付加されたパターンである。例えば、「A B C X」または「X Y A」などのパターンは、独立してそれ自身何らかの処理に関連付いているわけではないので、意味のないパターンである。意味の無いパターンの他の類型は、例えば、ある処理に関連するパターンと他の処理に関連するパターンとが時系列的に混じっているパターンである。例えば、「A B X C Y」または「X A Y B C」などのパターンは、独立してそれ自身何らかの処理に関連付いているわけではないので、意味のないパターンである。

[0054] 選択部 1 3 0 の詳細を説明する。選択部 1 3 0 は、抽出部 1 2 0 が抽出し

た複数のパターンのうちから、「意味のあるパターン」を選択し、選択したパターンを「一連のメッセージを処理単位でグループ化する情報」として出力する。

[0055] 上述した（前提３）を鑑みると、ある処理に関連するパターンと他の処理に関連するパターンとが時系列的に混じっているパターンは稀である。そのため、選択部１３０は、所定の閾値よりも低い頻度のパターンを選択しないことによって、このようなパターンをノイズとして除去することができる。

[0056] 同様に、上述した（前提３）を鑑みると、情報処理システム９００が実行する特定の処理に関連付くパターン「ＡＢＣ」および「ＸＹ」が抽出される頻度は、他のパターンの頻度よりも高くなると考えられる。しかし、例えば、パターン「ＡＢＣ」の頻度が高い場合には、当該パターンの部分集合であるパターン「ＡＢ」または「ＢＣ」の頻度も、当然に高くなる。また、当該パターンに何らかのイベントが付加された「ＡＢＣＸ」または「ＹＡＢＣ」などのパターンの頻度もある程度高くなることが考えられる。よって選択部１３０は、単純に頻度の高いパターンを選択するのみでは、「意味のあるパターン」を選択することはできない。

[0057] なぜなら、選択部１３０が単純に頻度の高いパターンを抽出すると、「意味のあるパターン」、「意味のあるパターンの部分集合のパターン」または「意味のあるパターンに何らかのイベントが付加されたパターン」が区別なく抽出されるからである。選択部１３０はこれらのパターンから「意味のあるパターン」を区別して選択することはできない。

[0058] 選択部１３０は、以下に示す動作によって、これら複数のパターンから、「意味のあるパターン」を選択する。以下、選択部１３０の動作を具体的に説明する。

[0059] 選択部１３０は、あるパターンを構成する時系列的に連続する複数のイベントと他のパターンを構成する時系列的に連続した複数のイベントとの包含関係に基づいて、複数のパターンから出力するパターンを選択する。また、選択部１３０は、パターンに関連づけられた頻度情報に基づいて、複数のパ

ターンから出力するパターンを選択する。

[0060] 理解を容易にするため、「親パターン」、「子パターン」という用語を定義する。一方のパターンを構成する時系列的に連続する複数のイベントが、他方のパターンを構成する時系列的に連続する複数のイベントの部分集合である場合に、前記他方のパターンを前記一方のパターンの親パターンと定義する。また、前記一方のパターンを前記他方のパターンの子パターンと定義する。例えば、「A B C D」というパターンは、「A B C」というパターンの親パターンである。例えば、「A B」というパターンは、「A B C」というパターンの子パターンである。同様に、「孫パターン」という用語等も定義することができる。

[0061] 選択部 130 は、抽出部 120 が抽出したパターンのそれぞれについて親子関係を分析する。図 6 は、選択部 130 が「A B C D」というパターンについて親子関係を分析した様子を表す図である。図 6 に示すように、選択部 130 は、イベントの順序が共通するパターンを同じ分類集合にまとめる。そして、選択部 130 は、パターンに含まれるイベント数（要素数）に基づいて、パターンの親子関係を分析する。

[0062] 以下、選択部 130 の動作として、具体例 1 から具体例 4 までを説明する。選択部 130 は、例えば、下記に示す具体例を組み合わせることにより、「意味のあるパターン」を抽出する。

[0063] （具体例 1）

選択部 130 は、前記複数のパターンのそれぞれについて親子関係を分析する。選択部 130 による分析の結果、例えば、パターン「A B C X」に対して、親パターンが「A B C X」であり、子パターンが「A B C」であり、孫パターンが「A B」とであるというような親子関係が得られる。

[0064] 選択部 130 は、親パターンの頻度に対する子パターンの頻度の比を算出する。図 5 に示した例では、親パターン「A B C X」の頻度は 17 であり、子パターン「A B C」の頻度は 43 であるので、頻度の比は $43 / 17 = 2.5$ である。ここで「/」は除算を表す記号である。ここで、当該比が比較

的大きな値の場合（例えば2以上の場合など）、当該親パターンは「意味の無いパターン」である可能性が高いと考えられる。なぜなら、そのような親パターンは、子パターンに単にイベントが付加しただけのパターンと考えられるからである。

[0065] 同様に選択部130は、親パターンの頻度に対する子パターンの頻度の比を算出する。図5に示した例では、親パターン「ABC」の頻度は43であり、子パターン「AB」の頻度は46であるので、頻度の比は $46 / 43 = 1.07$ である。当該比が1に近い値である場合（例えば1.1以下の場合など）、当該子パターンは「意味の無いパターン」である可能性が高いと考えられる。なぜなら、このような子パターンのほとんどは親パターンの部分集合として抽出されているにすぎず、このような子パターンが独立して何らかの処理に関連付いているとは考えにくいためである。このような子パターンが、親パターンの部分集合として以外に抽出される場合としては、例えば、例えば図4に示すロググループ2と3との境界のように、分割部110が、たまたま、「意味のあるパターン」の途中で正常ログ910を分割してしまうような場合、と考えられる。

[0066] （具体例2）

選択部130は、前記複数のパターンのそれぞれについて親子関係を分析する。選択部130は、パターンを構成する要素（イベント）の数が多く、かつ、頻度の高いパターンを、優先的に「意味のあるパターン」として選択してもよい。例えば、パターン「ABC」は、パターンを構成する要素の数が3つ（AとBとC）であり、頻度が43と高いので、優先的に「意味のあるパターン」として選択される。パターンを構成する要素の数の多少の判断基準、および、頻度の高低の判断基準は、例えば、あらかじめ定められていてもよい。

[0067] （具体例3）

選択部130は、前記複数のパターンのそれぞれについて親子関係を分析する。図7は、選択部130が、各パターンについて、当該パターンが幾つ

の親パターンを有するかを分析した表である。選択部 130 は、よりたくさん
の親パターンを有する子パターンを、優先的に「意味のあるパターン」と
して選択してもよい。

[0068] (具体例 4)

選択部 130 は、前記複数のパターンのそれぞれについて親子関係を分析
する。選択部 130 は、パターンの支持度、子パターンに対する親パター
ンの信頼度、または、支持度に対する信頼度の比（リフト値）等に基づいて、
「意味のあるパターン」を選択してもよい。支持度、信頼度、またはリフト
値は、例えばあらかじめオペレータにより定義されていてもよい。

[0069] (具体例 5)

選択部 130 は、例えば、「X Y A B C」というパターンが抽出された場
合に、当該パターンを充足するような子パターンを優先的に選択してもよい。
例えば、上述の例では、「A B C」と「X Y」という子パターンは、「A
B C X Y」という親パターンを充足するので、優先的に選択される。例えば、
上述の例では、「X A B」や「X B C」という子パターンでは、「A B C
X Y」という親パターンを充足しない（例えば、X A B では Y C が余る）の
で、候補から除外される。

[0070] 以上、選択部 130 の動作として（具体例 1）、（具体例 2）、（具体例
3）、（具体例 4）および（具体例 5）を説明した。選択部 130 は、上述
した具体例に示す動作により、抽出部 120 が抽出したパターンのうち、「
A B C」および「X Y」を、「意味のあるパターン」として選択する。選択
部 130 は、選択したパターンを、正常パターン抽出部 140 に出力する。

[0071] 正常パターン抽出部 140 は、選択部 130 から「意味のあるパターン」
の入力を受け付ける。正常パターン抽出部 140 は、当該パターンに基づい
て、正常ログ 910 から正常パターン 920 を抽出する。

[0072] 次に、ログ分析システム 100 の動作を説明する。図 8 は、第 1 の実施形
態にかかるログ分析システム 100 の動作を示すフローチャートである。

[0073] 分割部 110 は、正常ログ 910（ログファイル）を、複数のロググルー

プに分割する（ステップS101）。抽出部120は、複数のロググループからパターンを抽出する（ステップS102）。選択部130は、抽出部120が抽出したパターンのうちから、「意味があるパターン」を選択し、選択された「意味があるパターン」を出力する（ステップS103）。正常パターン抽出部140は、選択部130が選択したパターンに基づいて、正常ログ910から正常パターン920を抽出する（ステップS104）。

[0074] 次に、ログ分析システム100を実現可能なハードウェア構成の一例を説明する。図9はログ分析システム100を実現可能なハードウェア構成の一例を示す図である。

[0075] 本実施形態のログ分析システム100及び他の実施形態に係るログ分析システムは、図9に示すハードウェア構成を備えるコンピュータによって実現できる。そのコンピュータは、CPU（Central Processing Unit）1、メモリ2、記憶装置3、通信インターフェース（I/F）4を備える。ログ分析システムは、入力装置5または出力装置6を備えていてもよい。ログ分析システムの機能は、例えばCPU1が、メモリ2に読み出されたコンピュータプログラム（ソフトウェアプログラム、以下単に「プログラム」と記載する）を実行することにより実現される。実行に際して、CPU1は、通信インターフェース4、入力装置5および出力装置6を適宜制御する。

[0076] 尚、本実施形態および後述する各実施形態を例として説明される本発明は、係るプログラムが格納されたコンパクトディスク等の不揮発性の記憶媒体8によっても構成される。記憶媒体8が格納するプログラムは、例えばドライブ装置7により読み出される。

[0077] ログ分析システムが実行する通信は、例えばOS（Operating System）が提供する機能を使ってアプリケーションプログラムが通信インターフェース4を制御することによって実現される。入力装置5は、例えばキーボード、マウスまたはタッチパネルである。出力装置6は、例えばディスプレイである。ログ分析システムは、2つ以上の物理的に分離した装置が有線または無線で通信可能に接続されることによって構成されていてもよい。なお、ログ

分析システムおよびその各機能ブロックのハードウェア構成は、上述した構成に限定されない。また、上述したハードウェア構成は、後述する他の実施形態にかかるログ分析システム、および、障害原因分析システムにも適用可能である。

[0078] 第1の実施形態にかかるログ分析システム100の効果を説明する。ログ分析システム100は、一連のメッセージを処理単位でグループ化する情報が事前に与えられることなしに、正常ログ910から正常パターン920を適切に抽出すること、または抽出することを支援することができる。その理由は、ログ分析システム100は、正常ログ910に基づいて「一連のメッセージを処理単位でグループ化する情報」を自ら抽出することができるからである。

[0079] (第1の実施形態の変形例)

分割部110と、抽出部120と、選択部130と、正常パターン抽出部140は、それぞれ、例えば、図示しない入力装置からユーザあるいは他のプログラムやソフトウェアなどによるパターン生成開始の命令を受信することを契機に動作を開始してもよい。分割部110と、抽出部120と、選択部130と、正常パターン抽出部140は、それぞれ、例えば、ログファイルの入力、更新などを契機に動作を開始してもよい。

[0080] 分割部110は、ログファイルを、所定の個数のログレコードが1つのロググループを構成するように分割してもよい。例えば、分割部110は、ログファイルを、5つのログレコードが1つのロググループを構成するように、分割してもよい。分割部110は、所定の時間間隔でログファイルを取得してもよい。分割部110は、例えば、5分間分のログファイルを5分間隔で取得する、5分間分のログを3分間隔で取得する、というように動作してもよい。

[0081] 抽出部120は、時系列的に連続しないイベントをパターンと見なして抽出してもよい。例えば抽出部120は、ロググループ1「ABCXY」から、パターン「ABXY」、「ACY」または「BCY」などのパターンを抽

出してもよい。

[0082] 選択部 130 は、時系列的に連続しないイベントに基づいて親子関係を分析してもよい。例えば、選択部 130 は、親パターン「ABCX」に対して、「AC」または「ACX」などのパターンを子パターンとして関連づけてもよい。

[0083] <第 2 の実施形態>

図 10 は、本発明の第 2 の実施形態にかかるログ分析システム 101 の構成を示すブロック図である。図 3 に示した構成と実質的に同一の構成については、同様の符号を付与し、説明を省略する。第 2 の実施形態にかかるログ分析システム 101 は、第 1 の実施形態にかかるログ分析システム 100 に加えて、更に、変換部 150 を備える。

[0084] 第 1 の実施形態において、ログ分析システム 100 が受け付ける正常ログ 910 は、図 2 に示すように、ログレコード毎に既にイベント ID が付与されていた。第 2 の実施形態において、ログ分析システム 101 が受け付ける正常ログ 911 は、より生データに近いログファイルである。変換部 150 は、例えば、図 11A、図 11B、又は図 11C に示すようなログファイルを受け取り、受け取ったログファイルを図 2 に示すようなログに変換する。

[0085] 図 11A、図 11B および図 11C は、それぞれ、変換部 150 が受け取る正常ログ 911（ログファイル）の一例を示す図である。ログファイルは、少なくとも 1 以上のログレコードで構成される。ログレコードは、少なくとも 1 以上のログ要素を含んで構成される。ログ要素とは、日時、ログ ID（識別子）、メッセージなどログレコードを構成する要素である。

[0086] 変換部 150 は、事前にオペレータにより定義されたフォーマットに従って、正常ログ 911 を変換する。図 12 は、オペレータが事前に定義したフォーマットの一例を表す図である。図 12 に示すフォーマットは、「2013/07/01 09:00:00～2013/07/01 18:00:00」という期間を指定し、「日時」「ログ ID」というログ要素を指定している。なお、フォーマットにはログファイルからログ要素を抽出するために、各ログ要素の正規表現などの、抽出する

ログ要素を識別可能な情報を記録してもよい。また、フォーマットには、対象とするログファイルの名称など、ログファイルを識別可能な情報を記録してもよい。フォーマットは、ログ要素により構成されるログレコードの構造を正規表現として記録してもよい。

[0087] 図13は、オペレータが事前に定義したフォーマットの他の例を表す図である。図13は、イベントIDと、当該イベントIDが付与されるべきイベントの表現（レコード表現）とを関連づける情報を表す。レコード表現は、各ログレコードを表す文字列である。レコード表現として、各ログレコードがそのまま記録されていてもよい。また、レコード表現として、各ログレコードのログ要素を抽出した情報が記録されていてもよい。あるいは、レコード表現として、各ログレコードと合致する正規表現が記録されていてもよい。

[0088] 変換部150は、例えば図11A、図11B、又は図11Cに示すような正常ログ911の入力を受け付けると、図12または図13に示すフォーマットに基づいて、図2に示すようなログファイルに変換する。変換部150は、例えば、ログフォーマットにおいて指定された期間分のログを情報処理システム900から読み込んでもよい。なお、読み込む対象のログファイルは複数存在していてもよい。変換部150は、取得した、または、読み込んだログファイルから、フォーマットが指定するログ要素を抽出する。例えば、図12に示すフォーマットの場合、変換部150は、ログファイルに含まれるログレコードから、日時と、ログIDのみを抽出する。変換部150は、変換したログを時系列順に並べ替え、出力する。これにより、分割部110、抽出部120、選択部130、正常パターン抽出部140は、正常ログ911を処理可能となる。

[0089] 図14は、第2の実施形態にかかるログ分析システム101の動作を説明するフローチャートである。図8において説明したステップと同一の動作を行うステップに関しては、同一の符号を付与し、説明を省略する。

[0090] 変換部150は、ログファイル911を受け付け、受け付けたログファイ

ルを、各ログレコードにイベントIDが付与されるように変換する（ステップS201）。分割部110は、変換部150から変換されたログファイルを受け付ける。以降の動作は、図3におけるステップS101ないしステップS104に示した動作と同様であるので、説明を省略する。

[0091] 第2の実施形態にかかるログ分析システム101の効果を説明する。ログ分析システム101は、変換部150を備える。変換部150は、情報処理システム900が出力するログを、分割部110、抽出部120、選択部130、及び正常パターン抽出部140が処理可能な形式に変換する。これにより、ログ分析システム101は、様々なフォーマットで出力される正常ログ911を処理することができる。

[0092] <第3の実施形態>

図15は、本発明の第3の実施形態における障害原因分析システム300の構成を説明するブロック図である。障害原因分析システム300は、第2の変換部310と、パターン比較部320と、異常パターン出力部330とを備える。

[0093] 第2の変換部310は、情報処理システム900が出力した、分析対象のログファイルである分析対象ログ990を取得する。第2の変換部310には、あらかじめ、選択部130が選択した「一連のメッセージを処理単位でグループ化する情報」が登録されている。第2の変換部310は、分析対象ログ990から、「一連のメッセージを処理単位でグループ化する情報」に基づいて、パターンを抽出する。

[0094] パターン比較部320は、第2の変換部310が分析対象ログ990を変換したパターンと、正常パターン記憶部200が記憶するパターンとを比較する。パターン比較部320は、分析対象ログ990を変換したパターンにおいて、正常パターン920と異なるパターンを抽出する。

[0095] 異常パターン出力部330は、パターン比較部320が実行した比較の結果に基づいて、分析対象ログ990のうち、異常な箇所を出力する。

[0096] 第3の実施形態にかかる障害原因分析システム300によれば、情報処理

システム 900 が出力する膨大な量のログから、当該障害が発生した直接的または間接的な原因を示すメッセージを抽出し、出力することができる。

[0097] <第 4 の実施形態>

図 16 は、本発明の第 4 の実施形態にかかるログ分析システム 102 の構成を表すブロック図である。図 16 に示すように、ログ分析システム 102 は、分割部 112 と、抽出部 122 と、選択部 132 とを備える。

[0098] 分割部 112 は、ログファイルを、所定のルールに従って、複数のロググループへと分割する。ログファイルは、イベントと当該イベントが発生した時刻とが関連づけられた情報である複数のログレコードが時系列順に整列された情報を含む。ロググループは、時系列順に連続する複数のログレコードの集合である。

[0099] 抽出部 122 は、複数の前記ロググループから、時系列的に連続する複数のイベントによって構成される複数のパターンを抽出し、前記パターンを、当該パターンが幾つのログレコードから抽出されたかを表す頻度情報と関連づける。

[0100] 選択部 132 は、前記複数のパターンから出力すべきパターンを選択する。その際、選択部 132 は、あるパターンを構成する時系列的に連続する複数のイベントと、他のパターンを構成する時系列的に連続した複数のイベントとの包含関係に基づいて、かつ、前記パターンに関連づけられた頻度情報に基づいて、選択する。

[0101] また、上述した各実施の形態は、適宜組み合わせで実施されることが可能である。また、本発明は、上述した各実施の形態に限定されず、様々な態様で実施されることが可能である。

[0102] 各ブロック図に示したブロック分けは、説明の便宜上から表された構成である。各実施形態を例に説明された本発明は、その実装に際して、各ブロック図に示した構成には限定されない。

[0103] 以上、本発明を実施するための形態について説明したが、上記実施の形態は本発明の理解を容易にするためのものであり、本発明を限定して解釈する

ためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得ると共に、本発明にはその等価物も含まれる。

[0104] この出願は、2013年9月24日に提出された日本出願特願2013-197416を基礎とする優先権を主張し、その開示の全てをここに取り込む。

符号の説明

- [0105]
- | | |
|-----|------------|
| 1 | CPU |
| 2 | メモリ |
| 3 | 記憶装置 |
| 4 | 通信インターフェース |
| 5 | 入力装置 |
| 6 | 出力装置 |
| 7 | ドライブ装置 |
| 8 | 記憶媒体 |
| 100 | ログ分析システム |
| 101 | ログ分析システム |
| 102 | ログ分析システム |
| 110 | 分割部 |
| 112 | 分割部 |
| 120 | 抽出部 |
| 122 | 抽出部 |
| 130 | 選択部 |
| 132 | 選択部 |
| 140 | 正常パターン抽出部 |
| 150 | 変換部 |
| 200 | 正常パターン記憶部 |
| 300 | 障害原因分析システム |
| 310 | 第2の変換部 |

3 2 0	パターン比較部
3 3 0	異常パターン出力部
9 0 0	情報処理システム
9 1 0	正常ログ
9 1 1	正常ログ
9 2 0	正常パターン
9 9 0	分析対象ログ

請求の範囲

- [請求項1] イベントと当該イベントが発生した時刻とが関連づけられた情報である複数のログレコードが時系列順に整列された情報を含むログファイルを、所定のルールに従って、時系列順に連続する複数のログレコードの集合である複数のロググループへと分割する分割手段と、
- 複数の前記ロググループから、時系列的に連続する複数のイベントによって構成される複数のパターンを抽出し、前記パターンを、当該パターンが幾つのログレコードから抽出されたかを表す頻度情報と関連づける抽出手段と、
- あるパターンを構成する時系列的に連続する複数のイベントと、他のパターンを構成する時系列的に連続した複数のイベントとの包含関係に基づいて、かつ、前記パターンに関連づけられた頻度情報に基づいて、前記複数のパターンから出力すべきパターンを選択する選択手段と、
- を備えるログ分析システム。
- [請求項2] 前記ログファイルは、実行する処理に応じて複数のイベントを所定の順番で発生することが可能な情報処理装置から出力された情報に基づくログファイルであって、
- 前記選択手段は、特定の処理が実行されたことに応じて発生した一連のイベントを過不足なく含むパターンを選択する、
- 請求項1に記載のログ分析システム。
- [請求項3] 一方のパターンを構成する時系列的に連続する複数のイベントが、他方のパターンを構成する時系列的に連続する複数のイベントの部分集合である場合に、前記他方パターンを前記一方のパターンの親パターンと定義し、前記一方のパターンを前記他方のパターンの子パターンと定義する場合において、
- 前記選択手段は、前記複数のパターンを対象として親子関係を分析し、親パターンと子パターンとの間の、信頼度、支持度、またはリフ

ト値に基づいて、前記複数のパターンから出力するパターンを選択する、

請求項 2 に記載のログ分析システム。

[請求項 4]

一方のパターンを構成する時系列的に連続する複数のイベントが、他方のパターンを構成する時系列的に連続する複数のイベントの部分集合である場合に、前記他方パターンを前記一方のパターンの親パターンと定義し、前記一方のパターンを前記他方のパターンの子パターンと定義する場合において、

前記選択手段は、前記複数のパターンを対象として親子関係进行分析し、所定の閾値よりも大きい数量の親パターンを有する子パターンを優先的に選択する、

請求項 2 または 3 に記載のログ分析システム。

[請求項 5]

一方のパターンを構成する時系列的に連続する複数のイベントが、他方のパターンを構成する時系列的に連続する複数のイベントの部分集合である場合に、前記他方パターンを前記一方のパターンの親パターンと定義し、前記一方のパターンを前記他方のパターンの子パターンと定義する場合において、

前記選択手段は、前記複数のパターンのそれぞれについて親子関係进行分析し、あるパターンに対して親パターンと子パターンとの両方が存在する場合において、前記親パターンの頻度に対する前記あるパターンの頻度の割合が第 1 の閾値未満であって、かつ、前記あるパターンの頻度に対する前記子パターンの頻度の割合が第 2 の閾値以上である場合に、前記あるパターンを優先的に選択する、

請求項 2 から 4 のいずれかに記載のログ分析システム。

[請求項 6]

前記選択手段は、パターンを構成する時系列的に連続するイベントの数が所定の閾値よりも大きく、かつ、所定の閾値よりも高い頻度と関連づけられたパターンを優先的に選択する、請求項 2 から 5 のいずれかに記載のログ分析システム。

[請求項7] 前記ログファイルは、前記情報処理装置が正常に動作していた環境におけるログファイルであって、

前記選択手段が出力したパターンに基づいて、前記ログファイルから前記処理が実行された順序および頻度を示す情報である正常パターンを抽出する正常パターン抽出手段を更に備える、

請求項2から5のいずれかに記載のログ分析システム。

[請求項8] 前記選択手段が出力したパターンに基づいて、前記情報処理装置から出力された他のログファイルから、前記処理が実行された順序および頻度を示す情報を抽出し、前記抽出されたパターンと、請求項7に記載のログ分析システムが抽出した正常パターンと、を比較することにより、前記他のログファイルに異常なパターンが含まれているか否かを分析し、異常なパターンが含まれている場合には、当該異常なパターンを構成するログレコードを出力する、

障害原因分析システム。

[請求項9] コンピュータが、

イベントと当該イベントが発生した時刻とが関連づけられた情報である複数のログレコードが時系列順に整列された情報を含むログファイルを、所定のルールに従って、時系列順に連続する複数のログレコードの集合である複数のロググループへと分割し、

複数の前記ロググループから、時系列的に連続する複数のイベントによって構成される複数のパターンを抽出し、前記パターンを、当該パターンが幾つのログレコードから抽出されたかを表す頻度情報と関連づけ、

あるパターンを構成する時系列的に連続する複数のイベントと、他のパターンを構成する時系列的に連続した複数のイベントとの包含関係に基づいて、かつ、前記パターンに関連づけられた頻度情報に基づいて、前記複数のパターンから出力すべきパターンを選択する、

ログ分析方法。

[請求項10]

コンピュータに、

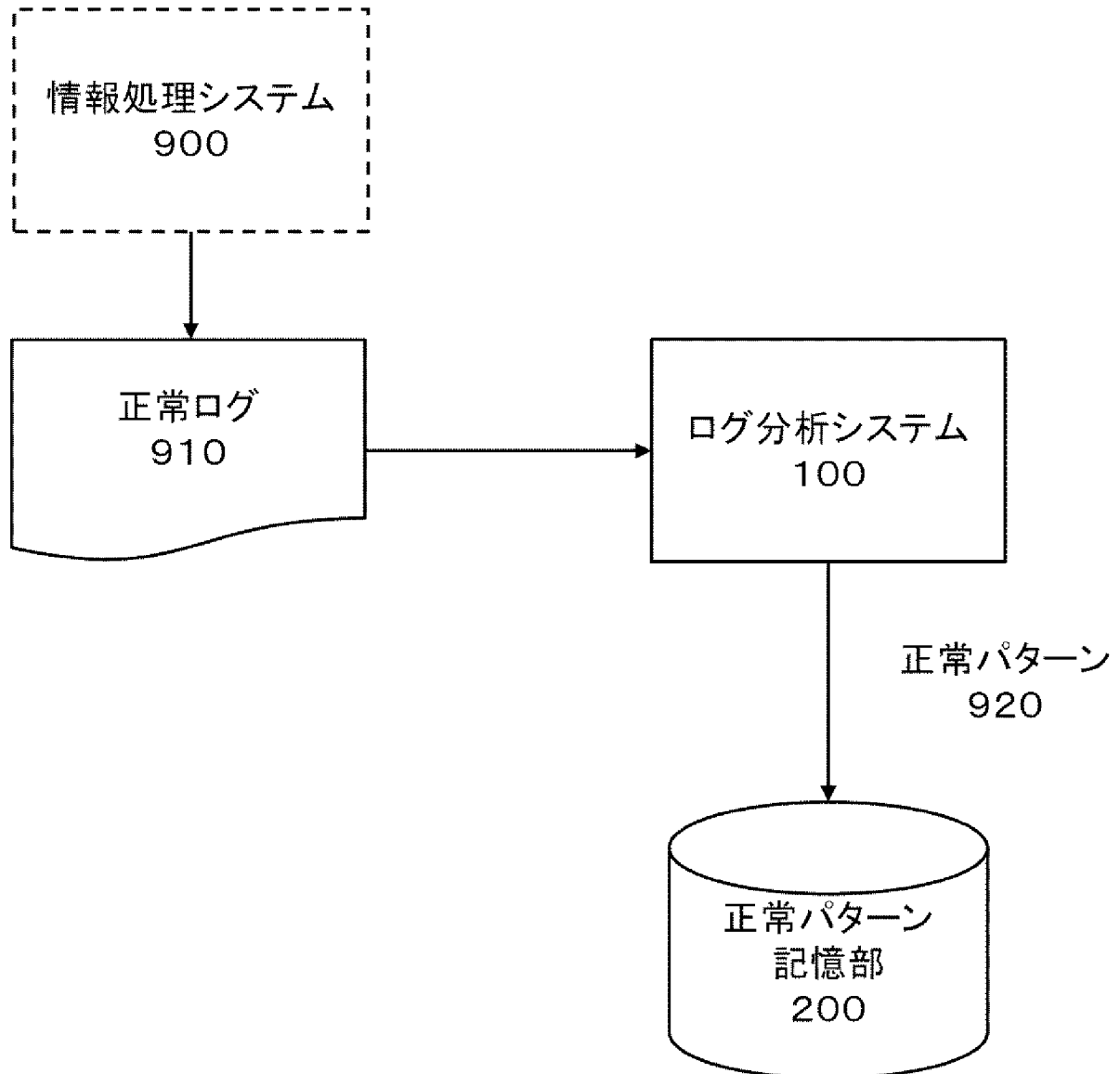
イベントと当該イベントが発生した時刻とが関連づけられた情報である複数のログレコードが時系列順に整列された情報を含むログファイルを、所定のルールに従って、時系列順に連続する複数のログレコードの集合である複数のロググループへと分割する処理と、

複数の前記ロググループから、時系列的に連続する複数のイベントによって構成される複数のパターンを抽出し、前記パターンを、当該パターンが幾つのログレコードから抽出されたかを表す頻度情報と関連づける処理と、

あるパターンを構成する時系列的に連続する複数のイベントと、他のパターンを構成する時系列的に連続した複数のイベントとの包含関係に基づいて、かつ、前記パターンに関連づけられた頻度情報に基づいて、前記複数のパターンから出力すべきパターンを選択する処理と、

を実行させるプログラムを記憶する記録媒体。

[図1]

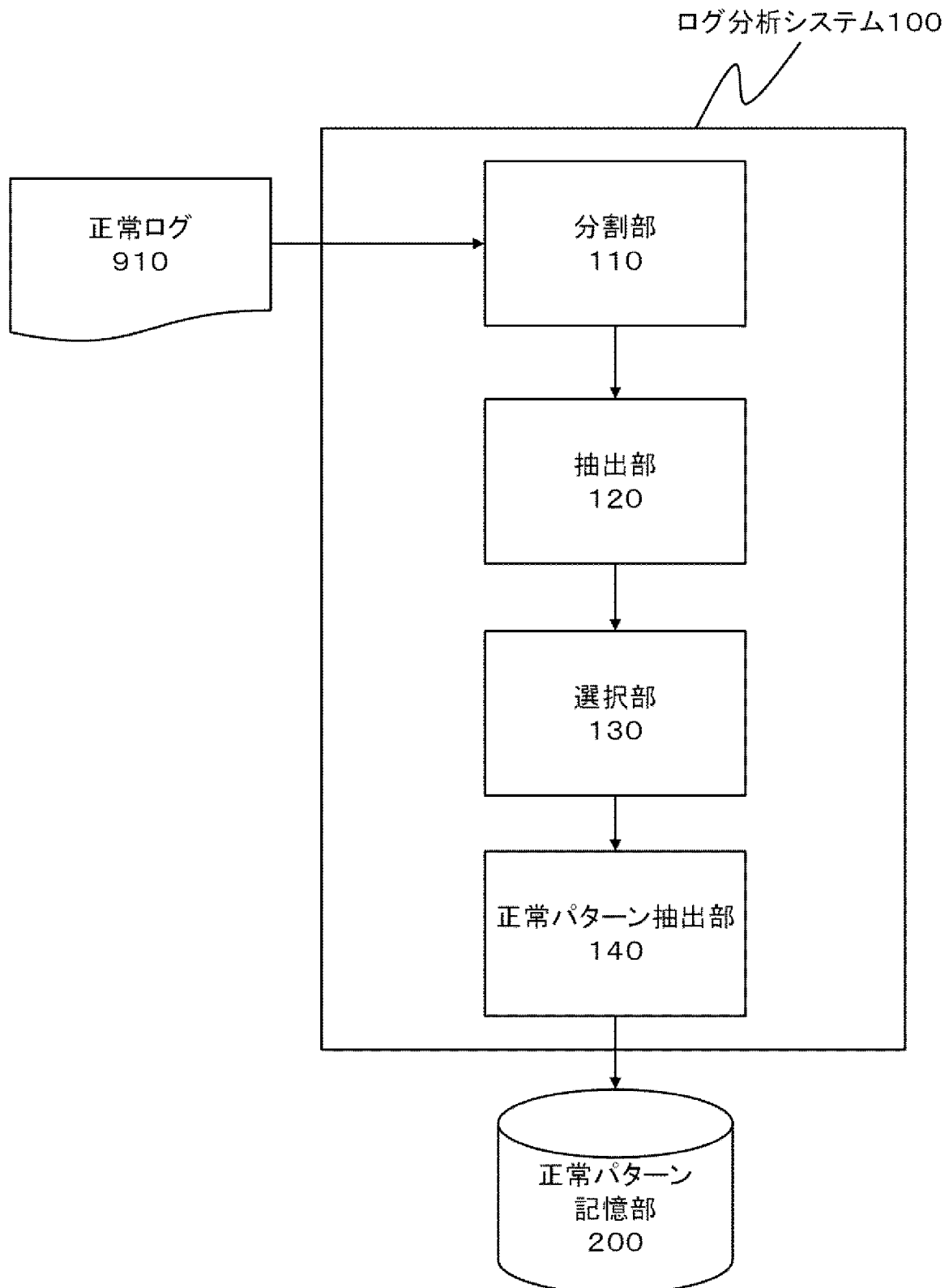


[図2]

正常ログ
910

イベントが発生した時刻	イベントID
09:04:01	A
09:04:02	B
09:04:03	C
09:04:04	X
09:04:05	Y
09:04:06	A
09:04:07	B
09:04:08	C
09:04:09	A
09:04:10	B
09:04:11	C
09:04:12	A
09:04:13	X
09:04:14	B
09:04:15	C
09:04:16	Y
...	...

[図3]



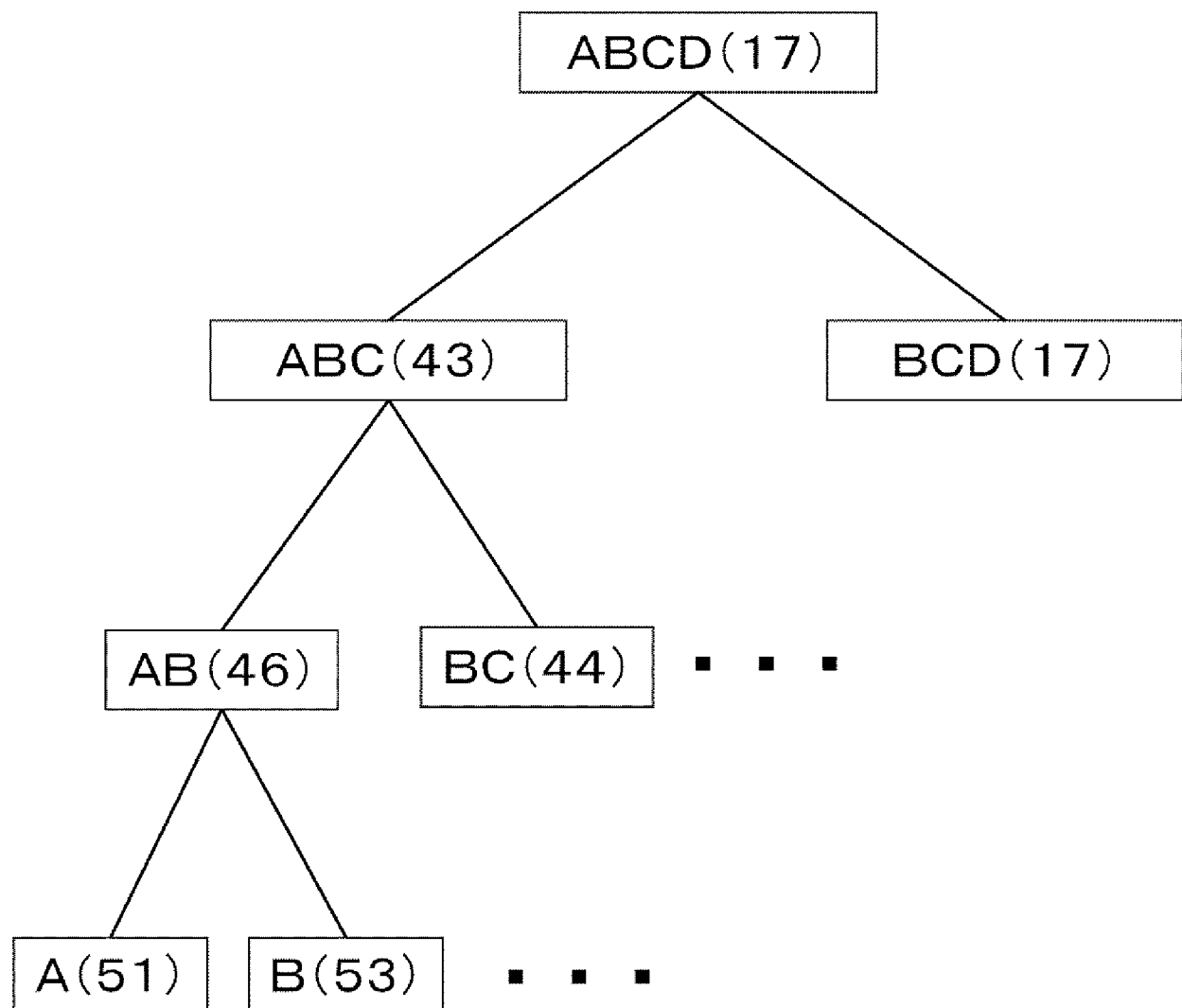
[図4]

時間情報	ロググループ	ログID
09:04:01～09:04:05	1	ABCXY
09:04:06～09:04:10	2	ABCAB
09:04:11～09:04:15	3	CAXBC
09:04:15～09:04:20	4	YABCX
09:04:21～09:04:25	5	YXYXY
...	...	...

[図5]

要素数	パターン	頻度
1	A	51
1	B	53
...	...	...
2	AB	46
2	XY	43
...	...	...
3	ABC	43
3	BCX	7
3	YAB	5
3	CXY	3
...	...	...
4	ABCX	17
4	BCXY	3
4	AXBC	1
...	...	...
5	ABCDE	1
...	...	...

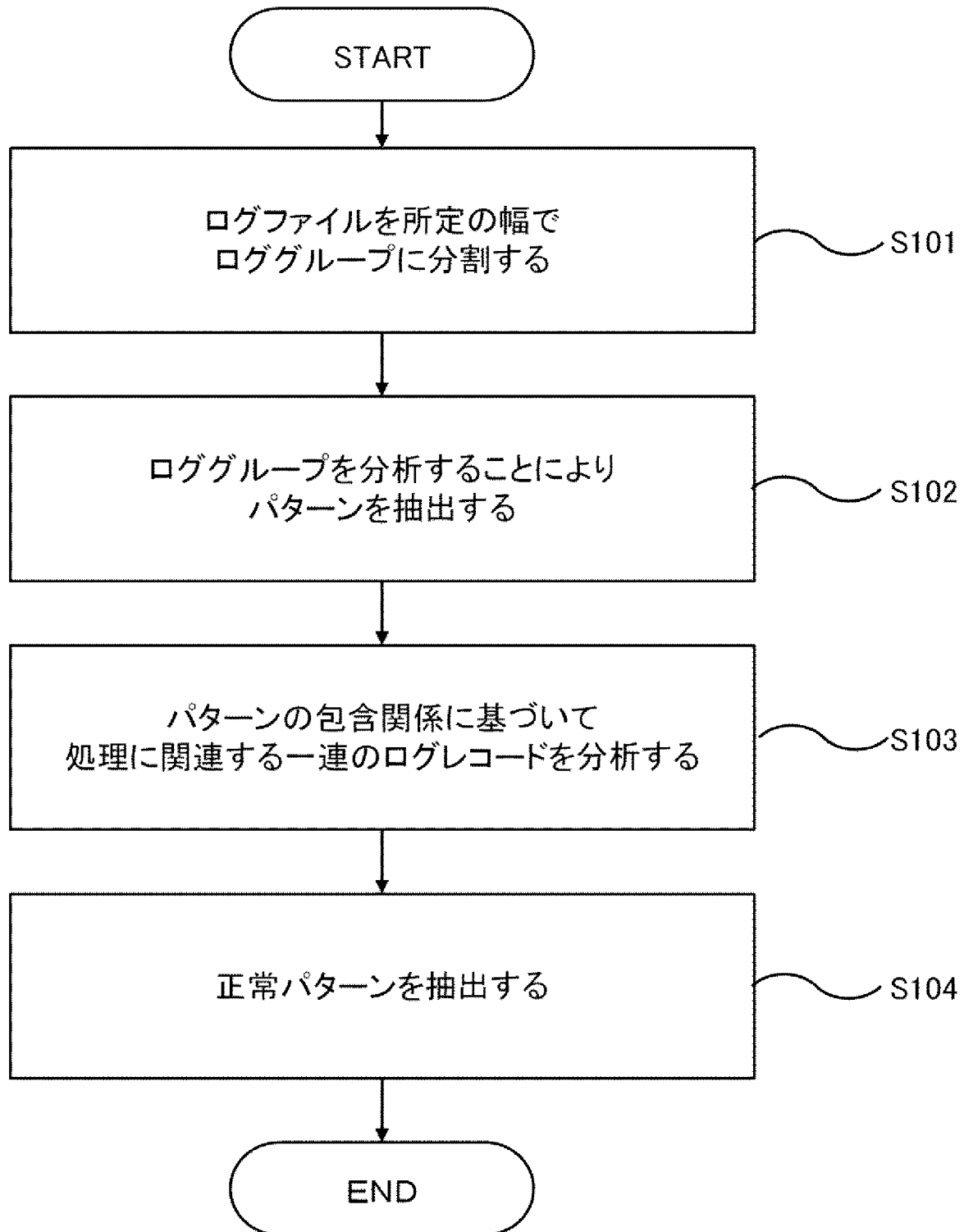
[図6]



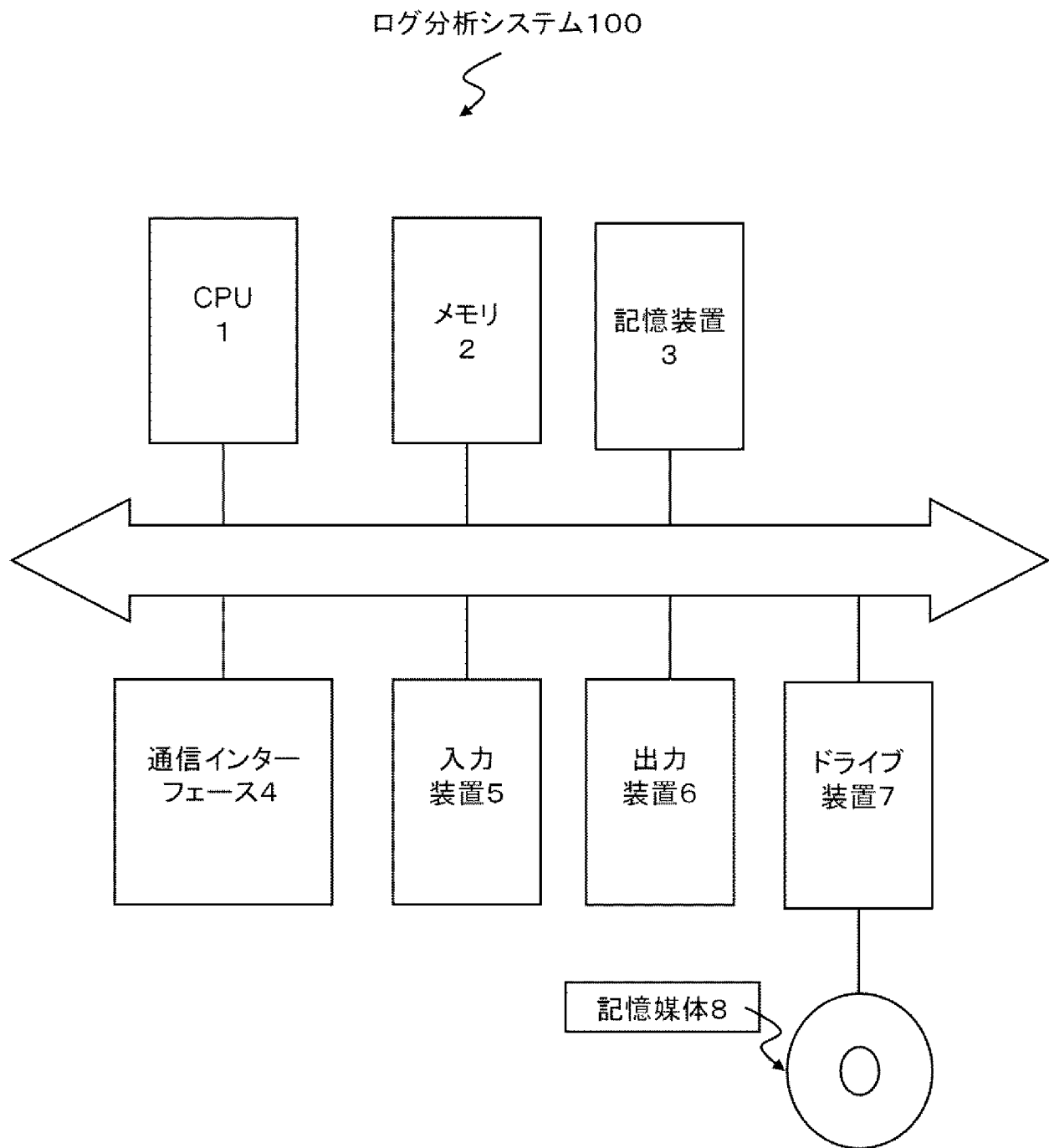
[図7]

子パターン	頻度	属する親の数
ABC	43	8
XY	39	6
XAB	29	2
XBC	27	2
...	...	...

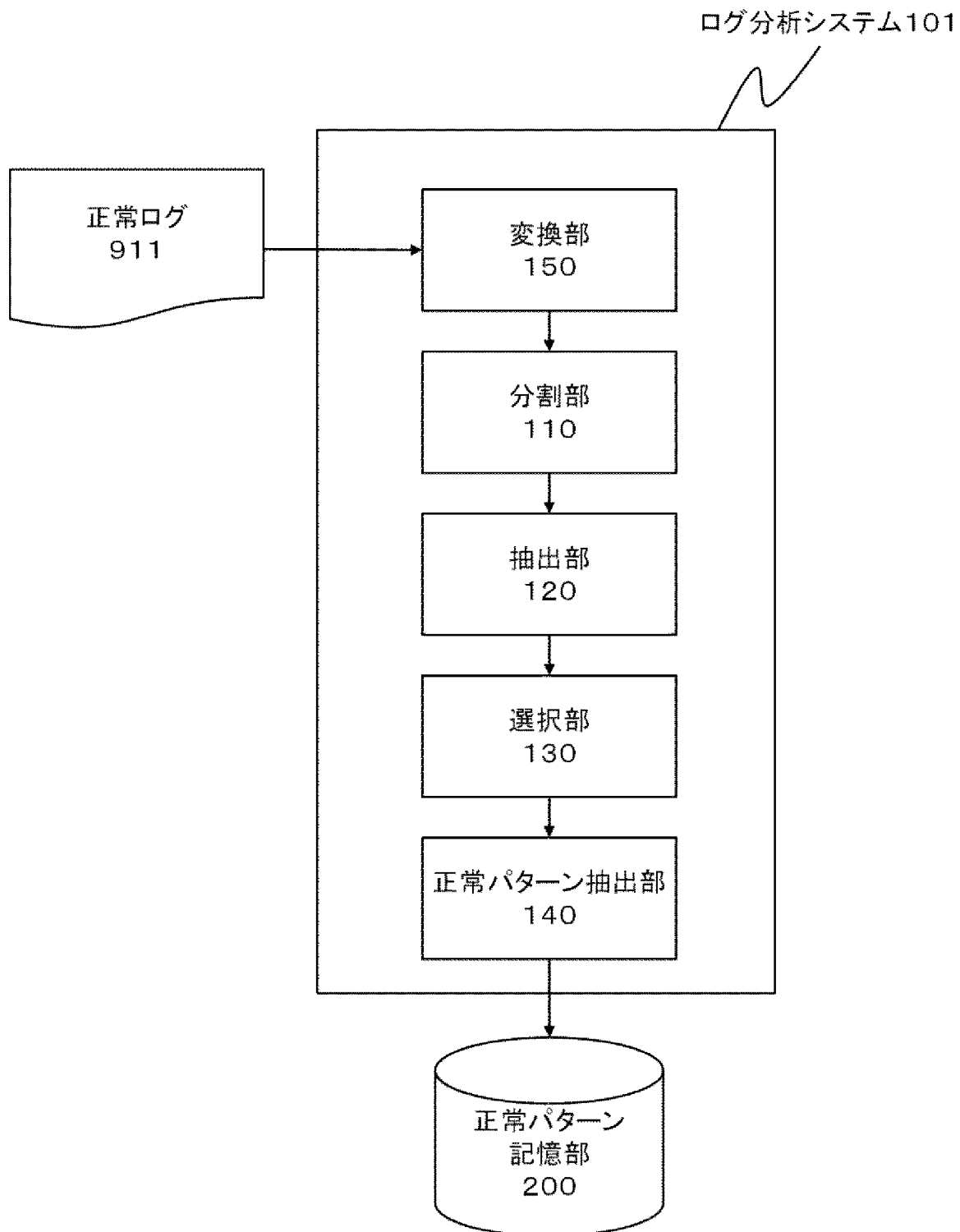
[図8]



[図9]



[図10]



[図11A]

```
mysqld.log
130201 09:04:01 mysqld started
130201 09:04:01 InnoDB : Started;
130201 09:04:01 [note] user/libexec/mysqld:
```

[図11B]

```

access.log
172.16.200.1 - - [2013/02/01 09:04:01] "GET /www/index.html
172.16.200.1 - - [2013/02/01 09:04:01] "GET /www/searchitems.html
172.16.200.3 - - [2013/02/01 09:05:03] "GET /www/buy.html

```

[図11C]

```

message
Feb 01 09:04:01 node1 httpd[1305] *****
Feb 01 09:04:03 node1 httpd[1298] *****

```

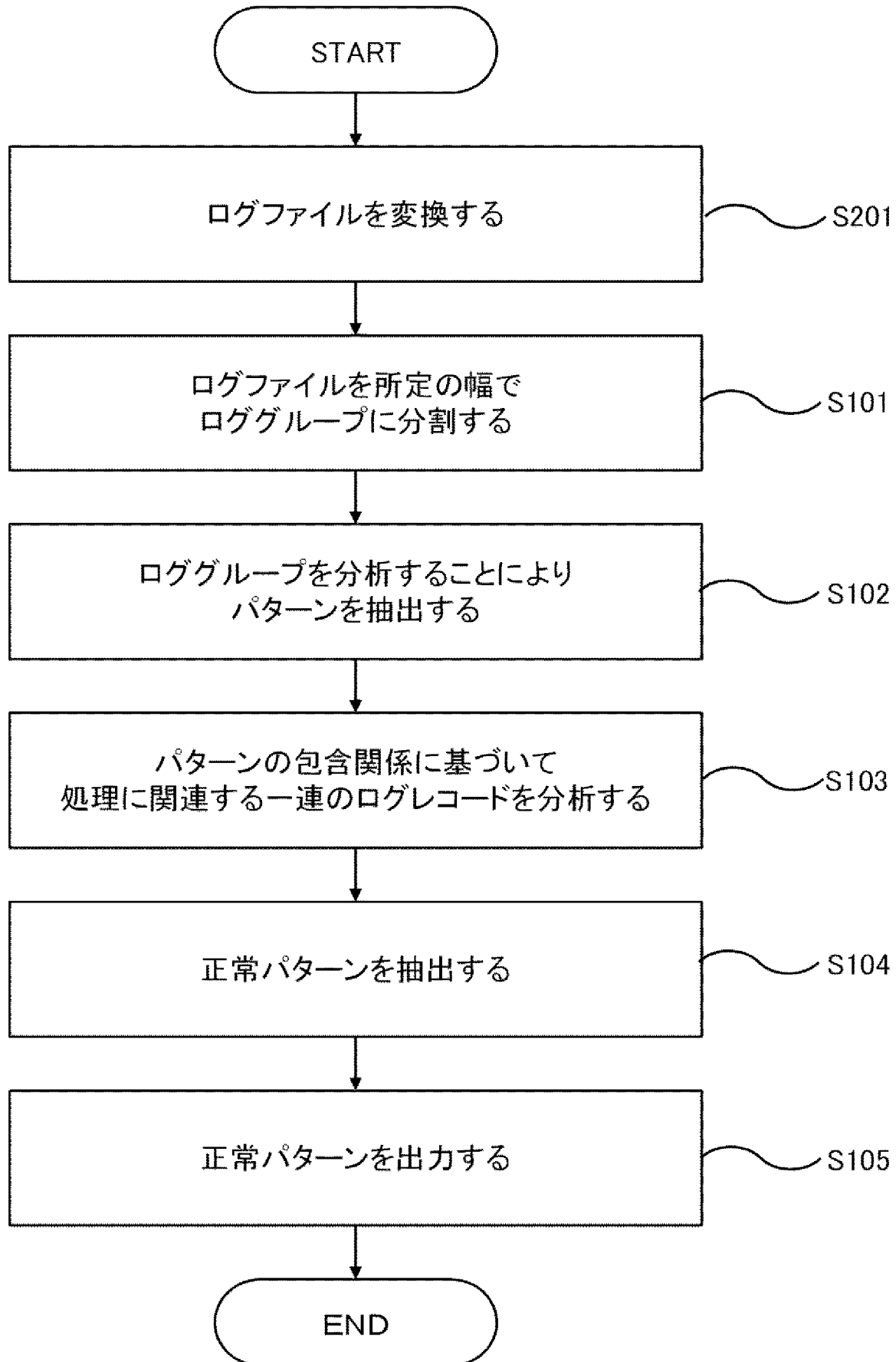
[図12]

分析期間	ログ要素
2013/07/01 09:00:00 ~ 2013/07/01 18:00:00	日時(yyyy/MM/DD, HH:mm:ss)、 ログID

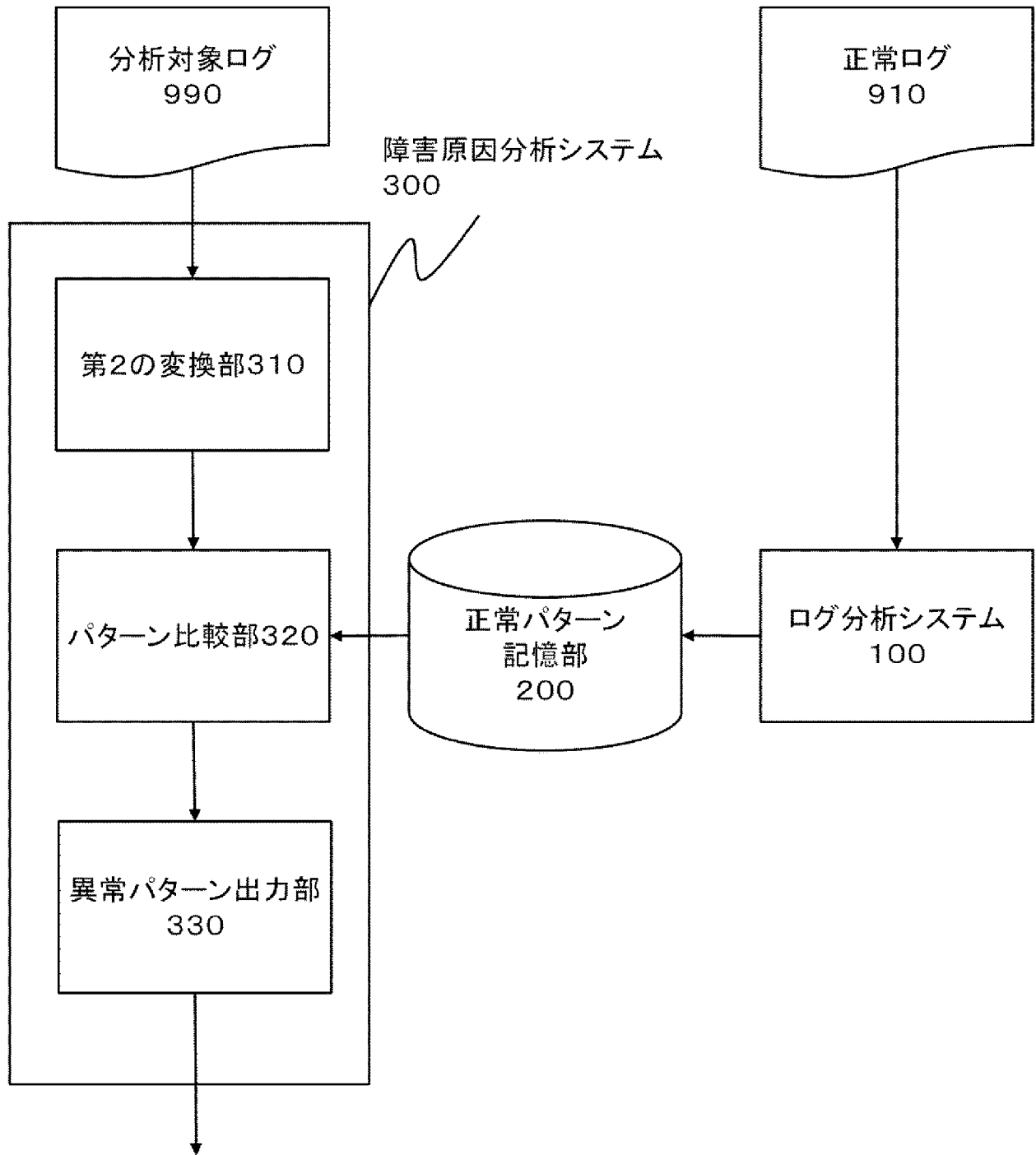
[図13]

ログID	レコード表現
A	mysqld started
B	[note] ***/***/***/mysqld failed
C	*** . *** . *** . *** stopped

[図14]

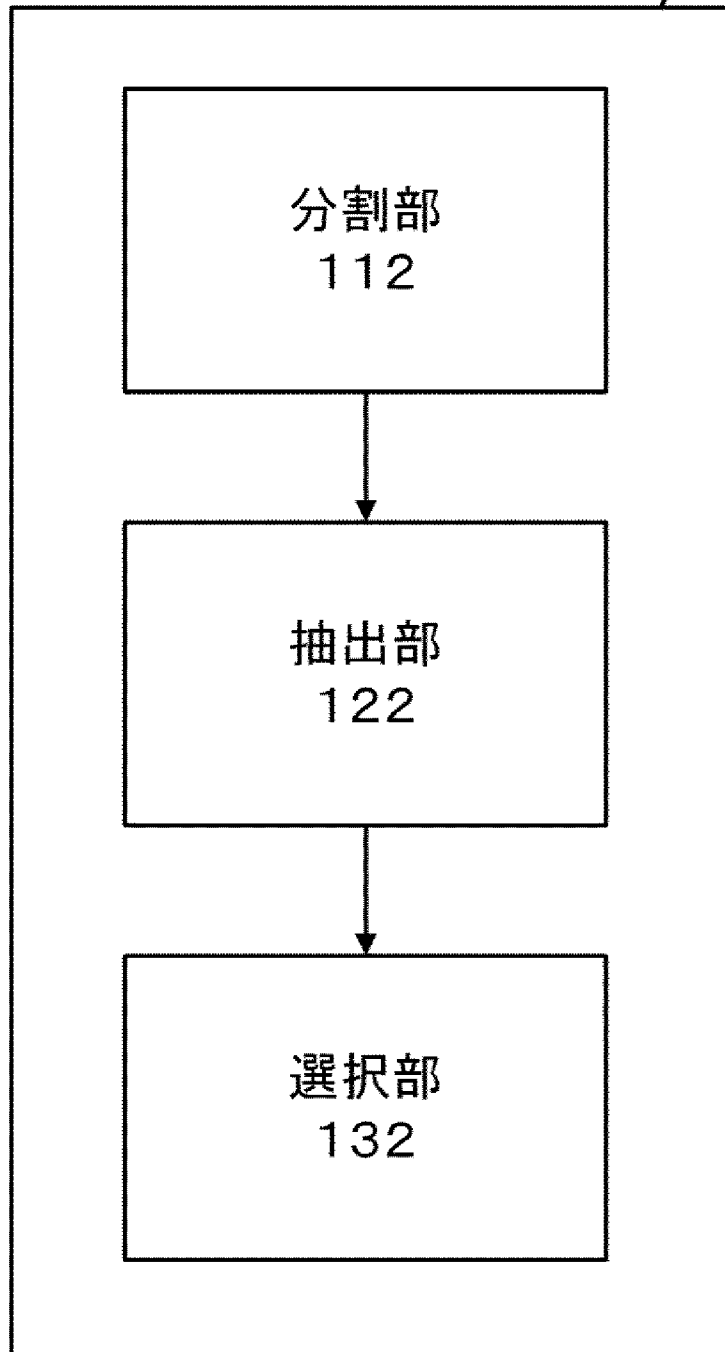


[図15]



[図16]

ログ分析システム102



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2014/004275

A. CLASSIFICATION OF SUBJECT MATTER

G06F11/34(2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F11/34

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	1996-2014
Kokai Jitsuyo Shinan Koho	1971-2014	Toroku Jitsuyo Shinan Koho	1994-2014

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 2012-123694 A (Fujitsu Ltd.), 28 June 2012 (28.06.2012), paragraphs [0009] to [0027]; fig. 1 to 3 & US 2012/0151282 A1	1-10
A	JP 2011-2870 A (Hitachi, Ltd.), 06 January 2011 (06.01.2011), paragraphs [0061] to [0064]; fig. 8 & US 2011/0047208 A1	1-10
A	JP 2006-318071 A (Fujitsu Ltd.), 24 November 2006 (24.11.2006), paragraphs [0013] to [0016], [0027] to [0042], [0067] to [0069]; fig. 5, 9 to 11 & US 2006/0256714 A1	1-10

☒ 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
28 October, 2014 (28.10.14)

Date of mailing of the international search report
04 November, 2014 (04.11.14)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2014/004275

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 11-224214 A (Fujitsu Ltd.), 17 August 1999 (17.08.1999), paragraphs [0030] to [0039]; fig. 8 (Family: none)	1-10
A	JP 2010-128661 A (Fujitsu Ltd.), 10 June 2010 (10.06.2010), paragraphs [0036] to [0051], [0070]; fig. 11 to 13 (Family: none)	1-10

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G06F11/34(2006.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（I P C））

Int.Cl. G06F11/34

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 4 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 4 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 4 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	JP 2012-123694 A（富士通株式会社）2012.06.28, 段落【0009】 - 【0027】, 第 1-3 図 & US 2012/0151282 A1	1-10
A	JP 2011-2870 A（株式会社日立製作所）2011.01.06, 段落【0061】 - 【0064】, 第 8 図 & US 2011/0047208 A1	1-10
A	JP 2006-318071 A（富士通株式会社）2006.11.24, 段落【0013】 - 【0016】, 【0027】 - 【0042】, 【0067】 - 【0069】, 第 5, 9-11 図 & US 2006/0256714 A1	1-10

☒ 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

2 8 . 1 0 . 2 0 1 4

国際調査報告の発送日

0 4 . 1 1 . 2 0 1 4

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）
郵便番号 1 0 0 - 8 9 1 5
東京都千代田区霞が関三丁目 4 番 3 号

特許庁審査官（権限のある職員）

大塚 俊範

5 B

4 6 8 0

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 4 5

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 11-224214 A (富士通株式会社) 1999. 08. 17, 段落【0030】-【0039】, 第 8 図 (ファミリーなし)	1-10
A	JP 2010-128661 A (富士通株式会社) 2010. 06. 10, 段落 【0036】 - 【0051】 , 【0070】 , 第 11-13 図 (ファミリーなし)	1-10