(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7514943号
(P7514943)

(45)発行日　令和6年7月11日(2024.7.11)

(24)登録日　令和6年7月3日(2024.7.3)

(51)国際特許分類
Ｇ０６Ｆ　11/34　(2006.01)

ＦＩ
Ｇ０６Ｆ　11/34　１７６

請求項の数　17　（全16頁）

(21)出願番号　特願2022-552777(P2022-552777)
(86)(22)出願日　令和3年3月3日(2021.3.3)
(65)公表番号　特表2023-516378(P2023-516378A)
(43)公表日　令和5年4月19日(2023.4.19)
(86)国際出願番号　PCT/US2021/020750
(87)国際公開番号　WO2021/178594
(87)国際公開日　令和3年9月10日(2021.9.10)
審査請求日　令和4年11月17日(2022.11.17)
(31)優先権主張番号　62/985,615
(32)優先日　令和2年3月5日(2020.3.5)
(33)優先権主張国・地域又は機関
米国(US)

(73)特許権者　508147326
シーメンス・ヘルスケア・ダイアグノスティックス・インコーポレイテッド
アメリカ合衆国ニューヨーク州１０５９１－５０９８．タリータウン．ベネディクトアベニュー５１１
(74)代理人　100127926
弁理士　結田 純次
(74)代理人　100140132
弁理士　竹林 則幸
(72)発明者　デュミットリィ・フラッドキン
アメリカ合衆国ペンシルベニア州１９０８７．ウェイン．ドンナ・レーン８
(72)発明者　タグバ・クラーチオグル
アメリカ合衆国ニュージャージー州０８
最終頁に続く

(54)【発明の名称】　複雑な物理的機器からのログの解析のための手法

(57)【特許請求の範囲】
【請求項１】
物理的機器によって生成されたログファイルを解析するためのコンピュータ実施方法であって：
物理的機器の１つまたはそれ以上の構成要素によって生成された１つまたはそれ以上のログファイルを受信することであって、各ログファイルは１つまたはそれ以上のログエントリを含むことと；
各ログファイルから、ログエントリの固定部分を表す複数のテンプレートを抽出することと；
ログファイル内のログエントリを複数のインスタンスにグループ化することであって、各インスタンスは、ログエントリにおけるデータを表す１つまたはそれ以上の次元に沿った複数の部分のうちの１つに対応することと；
インスタンスに含まれるテンプレートのセットを表す各インスタンスの表現を作成することと；
インスタンスの表現にクラスタリングプロセスを適用することによって複数のクラスタを生成することと；
グラフィカルユーザインタフェース（ＧＵＩ）において、クラスタおよびインスタンスの視覚表現を作成することと；
テンプレートおよびクラスタの組合せごとに、テンプレートがクラスタ内でどの程度一般的であるかを表す重要度値を計算することと；

クラスタがＧＵＩにおいてユーザによって選択されたとき、クラスタの重要度値を表示すること、を含み、
テンプレートおよびクラスタの重要度値は、テンプレートがクラスタ内にある第１の確率と、テンプレートが、ログファイルから抽出された複数のテンプレートにおける１つまたはそれ以上の他のテンプレート内にある第２の確率との差である、
前記方法。
【請求項２】
クラスタおよびインスタンスの視覚表現は、クラスタのプロットであり、該プロットはインスタンスを列挙するｘ軸を有する、請求項１に記載の方法。
【請求項３】
プロットは、時間値を列挙するｙ軸を有する、請求項２に記載の方法。
【請求項４】
グラフィカルユーザインタフェース内のクラスタのうちの１つが選択されたとき、クラスタに対応するログエントリのうちの１つまたはそれ以上を表示すること：
をさらに含む、請求項１に記載の方法。
【請求項５】
複数の部分は、ログエントリの生成に関連付けられた物理的機器の実行サイクルを表す時間次元を含む、請求項１に記載の方法。
【請求項６】
複数の部分は、ログエントリの生成に関連付けられた物理的機器内のロケーションを表す空間次元を含む、請求項１に記載の方法。
【請求項７】
複数の部分は、ログエントリの生成に関連付けられたジョブを表すジョブ次元を含む、請求項１に記載の方法。
【請求項８】
複数の部分は、ログエントリの生成に関連付けられたタスクを表すタスク次元を含む、請求項１に記載の方法。
【請求項９】
各インスタンスの表現は、テンプレートごとの要素を含むベクトルであり、該ベクトルの各要素は、特定のタイプのテンプレートのカウントまたはカウントの関数を記憶する、請求項１に記載の方法。
【請求項１０】
クラスタリングプロセスは：
クラスタリストを作成することと；
複数のインスタンスのうちの第１のインスタンスを第１のクラスタのプロトタイプとして宣言することと；
第１のクラスタをクラスタリストに追加することと；
複数のインスタンスにおける各追加のインスタンスについて、
クラスタリスト内のクラスタに対する追加のインスタンスの類似度測定値を計算することと、
類似度測定値について最高値を有するクラスタリスト内の特定のクラスタを識別することと、
最高値が所定の閾値よりも大きい場合、追加のインスタンスを特定のクラスタに追加することと、
最高値が所定の閾値以下である場合、クラスタリスト内に新たなクラスタを作成し、追加のインスタンスを新たなクラスタに追加することと、
を含む、請求項１に記載の方法。
【請求項１１】
類似度測定値はジャッカード類似度である、請求項１０に記載の方法。
【請求項１２】

物理的機器の１つまたはそれ以上の構成要素は、免疫測定法および臨床化学分析器システムの一部であり、各インスタンスは試験試料に対応する、請求項１に記載の方法。
【請求項１３】
物理的機器によって生成されたログファイルを解析するための製造物であって、方法を実行するためのコンピュータ実行可能命令を保持する非一時的有形コンピュータ可読媒体を含み、該方法は：
物理的機器の１つまたはそれ以上の構成要素によって生成された１つまたはそれ以上のログファイルを受信することであって、各ログファイルは１つまたはそれ以上のログエントリを含むことと；
各ログファイルから、ログエントリの固定部分を表す複数のテンプレートを抽出することと；
ログファイル内のログエントリを複数のインスタンスにグループ化することであって、各インスタンスは、ログエントリにおけるデータを表す複数の次元のうちの１つに対応することと；
インスタンスに含まれるテンプレートのセットを表す各インスタンスの表現を作成することと；
インスタンスの表現にクラスタリングプロセスを適用することによって複数のクラスタを生成することと；
ＧＵＩにおいて、クラスタおよびインスタンスの視覚表現を作成することと；
テンプレートおよびクラスタの組合せごとに、テンプレートがクラスタ内でどの程度一般的であるかを表す重要度値を計算することと；
クラスタがＧＵＩにおいてユーザによって選択されたとき、クラスタの重要度値を表示すること、を含み、
テンプレートおよびクラスタの重要度値は、テンプレートがクラスタ内にある第１の確率と、テンプレートが、ログファイルから抽出された複数のテンプレートにおける１つまたはそれ以上の他のテンプレート内にある第２の確率との差である、
前記製造物。
【請求項１４】
クラスタおよびインスタンスの視覚表現は、クラスタのプロットであり、該プロットはインスタンスを列挙するｘ軸を有する、請求項１３に記載の製造物。
【請求項１５】
方法は：
グラフィカルユーザインタフェース内のクラスタのうちの１つが選択されたとき、クラスタに対応するログエントリのうちの１つまたはそれ以上を表示することをさらに含む、請求項１３に記載の製造物。
【請求項１６】
物理的機器によって生成されたログファイルを解析するための製造物であって、方法を実行するためのコンピュータ実行可能命令を保持する非一時的有形コンピュータ可読媒体を含み、該方法は：
物理的機器の１つまたはそれ以上の構成要素によって生成された１つまたはそれ以上のログファイルを受信することであって、各ログファイルは１つまたはそれ以上のログエントリを含むことと；
各ログファイルから、ログエントリの固定部分を表す複数のテンプレートを抽出することと；
ログファイル内のログエントリを複数のインスタンスにグループ化することであって、各インスタンスは、ログエントリにおけるデータを表す複数の次元のうちの１つに対応することと；
インスタンスに含まれるテンプレートのセットを表す各インスタンスの表現を作成することと；
インスタンスの表現にクラスタリングプロセスを適用することによって複数のクラスタ

を生成することと；
ＧＵＩにおいて、クラスタおよびインスタンスの視覚表現を作成すること、を含み、
クラスタおよびインスタンスの視覚表現は、クラスタのプロットであり、該プロットはインスタンスを列挙するｘ軸を有する、
前記製造物。
【請求項１７】
物理的機器によって生成されたログファイルを解析するための製造物であって、方法を実行するためのコンピュータ実行可能命令を保持する非一時的有形コンピュータ可読媒体を含み、該方法は：
物理的機器の１つまたはそれ以上の構成要素によって生成された１つまたはそれ以上のログファイルを受信することであって、各ログファイルは１つまたはそれ以上のログエントリを含むことと；
各ログファイルから、ログエントリの固定部分を表す複数のテンプレートを抽出することと；
ログファイル内のログエントリを複数のインスタンスにグループ化することであって、各インスタンスは、ログエントリにおけるデータを表す複数の次元のうちの１つに対応することと；
インスタンスに含まれるテンプレートのセットを表す各インスタンスの表現を作成することと；
インスタンスの表現にクラスタリングプロセスを適用することによって複数のクラスタを生成することと；
ＧＵＩにおいて、クラスタおよびインスタンスの視覚表現を作成することと；
グラフィカルユーザインタフェース内のクラスタのうちの１つが選択されたとき、クラスタに対応するログエントリのうちの１つまたはそれ以上を表示すること、
を含む、前記製造物。
【発明の詳細な説明】
【技術分野】
【０００１】
関連出願の相互参照
本出願は、２０２０年３月５日に出願された「ＡＮ　ＡＰＰＲＯＡＣＨ　ＦＯＲ　ＡＮＡＬＹＳＩＳ　ＯＦ　ＬＯＧＳ　ＦＲＯＭ　Ａ　ＣＯＭＰＬＥＸ　ＰＨＹＳＩＣＡＬ　ＥＱＵＩＰＭＥＮＴ」と題する米国仮特許出願第６２／９８５，６１５号の利益を主張する。この仮特許出願の開示は、参照によってその全体が全ての目的で本明細書に組み入れられる。
【０００２】
本発明は、一般に、複雑な物理的機器によって生成されたログファイルの解析に関するシステム、方法および装置に関する。本明細書に記載の技法は、例えば、免疫測定法および化学分析器によって生成されるログエントリの解析に適用することができる。
【背景技術】
【０００３】
製造およびヘルスケアを含む多くの領域における技術開発が急速なペースで続いている。これらの領域で動作するシステムは、複数の構成をサポートし、最小限の人間の介入で多岐にわたるタスクを実行することが可能な高性能のマシンを、ますます用いている。例えば、免疫測定法および臨床化学分析を実行するためのシステムは、通常、各々が自律的にまたはほぼ自律的にタスクを実行する複数の多様なコンピューティングデバイスを有する。
【０００４】
複雑なシステムの動作中、その構成要素は、本明細書において、「ログデータ」または単に「ログ」と呼ばれるテキストファイルまたはバイナリファイルの形態のデータを生成する。これらのログは、機器によって実行されるソフトウェア内のプリントステートメントによって生成される。ログ内のデータは、機器がどのように動作しているかを理解する

のに有用な情報を含む。例えば、ログは、いくつかのメッセージが受信もしくは生成されたときのインジケーション、センサ測定値、受信したコマンドおよび実行された動作のシーケンス、所与の時点における変数の値、またはエラーが生じたときのシステムメモリのコンテンツを表すスタックトレースを含むことができる。

【０００５】

複雑なシステムは、入り組んだタスクを迅速に実行することが可能であるが、システムの複雑さにより、任意の課題を解決することが困難になる。従来、顧客はまず、システムに関与する問題を検出し、その問題を、システムの維持の任務を負う会社または組織に報告する。次に、領域専門家（サポートエンジニア等）がデータを手作業でレビューしなくてはならない。様々な要因に依拠して、問題の初期症状とその解決との間の時間が長くなり得る。解決速度は、サポートエンジニアの経験および問題の複雑さに依拠する。このトラブルシューティングプロセスは時間と労力がかかる。

【発明の概要】

【発明が解決しようとする課題】

【０００６】

本発明の実施形態は、動作を最適化し、問題を検出および診断するための、複雑な物理的機器からのログデータの自動解析に関する方法、システムおよび装置を提供することによって、上記の短所および欠点のうちの１つまたはそれ以上に対処し、克服する。この解析は、機器を用いて問題を迅速に識別する際に領域専門家をサポートすることができるか、または自律的または半自律的トラブルシューティングおよび修復手順に用いることができる。

【課題を解決するための手段】

【０００７】

いくつかの実施形態によれば、複雑な物理的機器によって生成されたログファイルを解析するためのコンピュータ実施方法が、物理的機器の１つまたはそれ以上の構成要素によって生成された１つまたはそれ以上のログファイルを受信することを含む。ログファイルの各々は、１つまたはそれ以上のログエントリを含む。各ログファイルから、ログエントリの固定部分を表す複数のテンプレートが抽出される。ログエントリは、ログファイルにおいて複数のインスタンスにグループ化される。各インスタンスは、ログエントリにおけるデータを表す１つまたはそれ以上の次元に沿った複数の部分のうちの１つに対応する。インスタンスに含まれるテンプレートのセットを表す各インスタンスの表現が作成される。インスタンスの表現にクラスタリングプロセスを適用することによって複数のクラスタが生成される。次に、グラフィカルユーザインタフェース（ＧＵＩ）において、クラスタおよびインスタンスの視覚表現を作成することができる。

【０００８】

他の実施形態では、複雑な物理的機器によって生成されたログファイルを解析するための製造物が、上記の方法を実行するためのコンピュータ実行可能命令を保持する非一時的有形コンピュータ可読媒体を含む。

【０００９】

他の実施形態によれば、複雑な物理的機器によって生成されたログファイルを解析するためのシステムが、非一時的有形コンピュータ可読媒体と、１つまたはそれ以上のプロセッサと、ディスプレイとを含む。コンピュータ可読媒体は、物理的機器の１つまたはそれ以上の構成要素によって生成された１つまたはそれ以上のログファイルを保持する。各ログファイルは、１つまたはそれ以上のログエントリを含む。プロセッサは、各ログファイルから、ログエントリの固定部分を表す複数のテンプレートを抽出し、ログファイル内のログエントリを複数のインスタンスにグループ化するように構成される。各インスタンスは、ログエントリにおけるデータを表す複数の次元のうちの１つに対応する。プロセッサは、インスタンスに含まれるテンプレートのセットを表す各インスタンスの表現を作成し、インスタンスの表現にクラスタリングプロセスを適用することによって複数のクラスタを生成するようにさらに構成される。ディスプレイは、ＧＵＩにおいて、クラスタおよび

インスタンスの視覚表現を提示する。
【００１０】
本発明のさらなる構成および利点は、添付図面を参照して進行する、以下の例示的な実施形態の詳細な説明から明らかとなろう。
【００１１】
本発明の上記態様および他の態様は、添付図面と併せて読むと、以下の詳細な説明から最もよく理解される。本発明を例示するために、現時点で好ましい実施形態が図示されているが、本発明は開示されている特定の手段に限定されないことを理解されたい。図面には以下の図が含まれる。
【図面の簡単な説明】
【００１２】
【図１】いくつかの実施形態による、ログデータの自動解析のためのシステムを示す。
【図２】本発明の異なる実施形態において適用することができるログ解析のための方法を示す。
【図３】ジャッカード類似度を用いたワンパスクラスタリングを実行する例示的な方法を示す。
【図４】本明細書において論じられる解析のために用いることができる例示的なグラフィカルユーザインタフェース（ＧＵＩ）を示す。
【図５】いくつかの実施形態において表示することができる試料特徴量重要度プロットを示す。
【図６】本発明の実施形態を実施することができる例示的なコンピューティング環境を示す。
【発明を実施するための形態】
【００１３】
本発明は、包括的には、動作を最適化し、問題を検出および診断するための、複雑な物理的機器からのログデータの自動解析に関する方法、システムおよび装置に関する。
【００１４】
図１は、いくつかの実施形態による、ログデータの自動解析のためのシステム１００を示す。この例において、構成要素１０５、１１０、１１５の群は、協働で、免疫測定法および臨床化学分析器システム１２０として動作する。当該技術分野において一般的に理解されているように、免疫測定法および臨床化学分析器システム１２０は、生体試料に対する試験を実行するために連続して動作する複数のコンピューティングデバイスを含む。免疫測定法および臨床化学分析器システム１２０は、本明細書に記載のログ解析技法を適用することができる複雑な物理的機器の１つの例であることを理解されたい。通常、これらの技法は、ログ解析データが生成される任意のタイプの複雑な物理的機器に適用することができる。
【００１５】
免疫測定法および臨床化学分析器システム１２０は、ネットワーク１３５を介してログファイルをログ処理コンピュータ１２５に転送する。このネットワーク１３５は、例えば、ローカルエリアネットワークまたはインターネットとすることができる。図１において、様々なデバイスをネットワーク１３５に連結する線が用いられているが、ネットワーク１３５は、有線または無線方式で動作し得ることを理解されたい。当該技術分野において一般的に既知の任意の技法を用いて、ネットワーク１３５を介したデータの転送を促進することができる。データ転送は、例えば、スケジューリングされた間隔で（例えば、毎時、毎晩等）、リアルタイムで（すなわち、ログファイルが生成される際に）、またはユーザからの要求時に行うことができる。
【００１６】
ログ処理コンピュータ１２５は、構成要素１０５、１１０および１１５によって生成されたログファイルを記憶する。図１には集中型のログ記憶装置が示されているが、他の実施形態では、異なる記憶アーキテクチャを利用することができることに留意されたい。例

えば、１つの実施形態では、各構成要素１０５、１１０および１１５は、自身のログファイルを、それぞれのユニット内にローカルで記憶する。他の実施形態では、免疫測定法および臨床化学分析器システム１２０は、構成要素１０５、１１０および１１５のそれぞれによって生成されたログファイルを記憶するローカル記憶機構を含む。ログ処理コンピュータ１２５は、図２および図３を参照して以下に説明する技法を用いて各ログファイルを解析する。この解析に基づいて、解析結果を表示するための１つまたはそれ以上のグラフィカルユーザインタフェース（ＧＵＩ）が生成される。次に、これらのＧＵＩを、（図１におけるような）ログ処理コンピュータ１２５に直接接続された、またはネットワーク１３５を介してログ処理コンピュータ１２５に間接的に接続されたユーザコンピュータ１３０上に表示することができる。

【００１７】

各ログファイルは、１つまたはそれ以上のログエントリを含む。本明細書に記載の技法は、各ログエントリがコンテンツ（すなわち、メッセージ）およびタイムスタンプを有すると仮定する。メッセージは、通常、構成要素１０５、１１０または１１５によって自動的に生成され、プログラムからのプリントステートメントに見られるように、固定部分（テンプレート）およびパラメータを含む。エントリの１つの例は以下の通りである：”Ｎｕｍｂｅｒｏｆ Ｔｉｐｓ Ａｖａｉｌａｂｌｅ： １７ ＮｕｍｂｅｒｏｆＴｉｐｓＣｏｍｉｔｔｅｄ： ０”ここで、「１７」および「０」はパラメータであり、残りはテンプレートである。

【００１８】

免疫測定法および臨床化学分析器システム１２０のソースコードに完全なアクセスを有する場合、出力ステートメントおよびその形式を識別し、ログエントリのいずれの部分がテンプレートであり、いずれがパラメータであるかを厳密に知ることができるであろう。上記のメッセージは、ｐｒｉｎｔ（”Ｎｕｍｂｅｒ ｏｆ Ｔｉｐｓ Ａｖａｉｌａｂｌｅ： {} ＮｕｍｂｅｒｏｆＴｉｐｓＣｏｍｉｔｔｅｄ： {} ”．ｆｏｒｍａｔ（１７．０））と同様のコードによって生成される可能性が高い。しかしながら、ソースコードへのアクセスを有することは稀にしか可能でなく、このため、自動的にログを編成し、テンプレートおよびパラメータを抽出する必要がある。

【００１９】

免疫測定法および臨床化学分析器システム１２０のいくらかの領域知識を与えられると、ログを自動的にパースして、メッセージ、ならびにインスタンスの開始および終了をマーキングするパラメータを識別することができる。例えば、領域知識は、メッセージ「Ｓｔａｒｔｉｎｇ ｃｙｃｌｅ ３６４」がサイクルの開始をマーキングすること、またはメッセージ「ＯｎＥｖｅｎｔＳａｍｐｌｅＡｒｒｉｖｅｄ（）． ＳａｍｐｌｅＵＩＤ： １３５１４」が、ｉｄ １３５１４を有するジョブの開始をマーキングすることを指定する場合がある。これは、（下記の）特定のインスタンスに属するメッセージを選択するのに十分である。さらに、当該技術分野において一般的に既知のテンプレート抽出方法を用いて、各メッセージをテンプレートにマッチングし、そのパラメータを識別することができる。例示的なテンプレート抽出方法は、Ｈｅ、Ｐｉｎｊｉａ、Ｊｉｅｍｉｎｇ Ｚｈｕ、Ｚｉｂｉｎ ＺｈｅｎｇおよびＭｉｃｈａｅｌ Ｒ． Ｌｙｕ．「Ｄｒａｉｎ： Ａｎ ｏｎｌｉｎｅ ｌｏｇ ｐａｒｓｉｎｇ ａｐｐｒｏａｃｈ ｗｉｔｈ ｆｉｘｅｄ ｄｅｐｔｈ ｔｒｅｅ．」Ｉｎ ２０１７ ＩＥＥＥ Ｉｎｔｅｒｎａｔｉｏｎａｌ Ｃｏｎｆｅｒｅｎｃｅ ｏｎ Ｗｅｂ Ｓｅｒｖｉｃｅｓ （ＩＣＷＳ）、 ｐｐ． ３３－４０． ＩＥＥＥ、 ２０１７ならびにＪｉｅｍｉｎｇ Ｚｈｕ、Ｓｈｉｌｉｎ Ｈｅ、Ｊｉｎｙａｎｇ Ｌｉｕ、Ｐｉｎｊｉａ Ｈｅ、Ｑｉ Ｘｉｅ、Ｚｉｂｉｎ Ｚｈｅｎｇ、Ｍｉｃｈａｅｌ Ｒ． Ｌｙｕ． Ｔｏｏｌｓ ａｎｄ Ｂｅｎｃｈｍａｒｋｓ ｆｏｒ Ａｕｔｏｍａｔｅｄ Ｌｏｇ Ｐａｒｓｉｎｇ． Ｉｎｔｅｒｎａｔｉｏｎａｌ Ｃｏｎｆｅｒｅｎｃｅ ｏｎ Ｓｏｆｔｗａｒｅ Ｅｎｇｉｎｅｅｒｉｎｇ （ＩＣＳＥ）、２０１９に記載されている。

【００２０】

ログファイルは、時間次元、空間次元およびジョブまたはタスクベースの次元を含むい

くつかの次元に沿って編成することができる。各次元は、解析のために、ログエントリを、本明細書において「インスタンス」と呼ばれる単一のデータ点にグループ化することを可能にする。ログファイルに関連付けられた様々なインスタンスを検討することにより、データエントリ内に提供される最も有用な情報の抽出が可能になる。インスタンスは互いに独立していないことを理解されたい。例えば、いくつかの実施形態では、解析は、異なるサイクル、ロケーションまたはジョブ間の依存性を検討することができる。

【００２１】

ログファイルの時間次元は、データがいつ生成されたか、または他の場合、どの時間が特定のコンテンツアイテムに関連しているかを表す。例えば、ログは、アクティビティのスケジューリングユニットであるサイクルに編成することができる。システムは、まずスケジューリングを行い、次に、特定のサイクルにおける或る数の動作の実行を試みる。サイクル識別子（ｉｄ）の値は、少なくともいくつかのメッセージに見ることができ、サイクルの開始／終了／ステータスを抽出することができる。このため、解析のために、単一のサイクルに対応するログのシーケンスを単一のインスタンスとして扱うことが可能である。空間次元は、データがどこで生成されるか、またはデータがどの構成要素もしくは機能を参照しているかを表す。システムは、大まかに、システム内の異なるロケーションとして考えることができる複数の別々の動作／機能を実行することができる。したがって、これらの「ロケーション」の各々を、別個のインスタンスとみなすことが可能である。最後に、ジョブまたはタスク次元は、ログエントリがどの機能を参照しているかを表す。システムは、１つまたはそれ以上のタスクを含む特定のジョブを与えられる。マシンが実行することができる個々のタスクの数は制限されているが、ジョブは、タスクと異なる入力との異なる組合せを含むことができる。別個のタスクは、システム内の別々のロケーションにも対応することができることにも留意されたい。

【００２２】

図２は、本発明の異なる実施形態において適用することができるログ解析のための方法２００を示す。この方法は、例えば、ログ処理コンピュータ１２５によって、または図１に示す構成要素１０５、１１０および１１５のうちの１つの中で実行することができる。

【００２３】

ステップ２０５から開始して、テンプレートは、システム構成要素から受信される１つまたはそれ以上のログメッセージから抽出される。上記で説明したように、各テンプレートは、インスタンスの開始および終了をマーキングするメッセージおよびパラメータを含む。抽出プロセスは、サンプリングおよび領域知識の組合せに基づいて導出することができる。より詳細には、長期間のログファイルは、領域専門家によって考察することができ、次に、専門家は、ログファイルがどのようにフォーマット設定されているかを示すことができる。例えば、１つの実施形態において、ログファイルは、数週間または数カ月の期間にわたって収集される。次に、単純なテキスト処理を適用し、ログファイル内の一意の行を識別することができる。行の一意性は、全ての文字、またはサブセットのみ（例えば、最初のｎ文字または最後のｎ文字、ここでｎはユーザが定義したパラメータである）に基づくことができる。次に、領域専門家は、いずれの文字が固定またはテンプレート部分であるか、およびいずれがパラメータに対応するかを示すことができる。当該技術分野において一般的に既知のテンプレート抽出方法をステップ２０５において利用することもできる。

【００２４】

抽出されたテンプレートを用いて、ステップ２１０において、ログが解析され、インスタンスにグループ化される。上述したように、これらのインスタンスは、空間的であるか、時間的であるか、またはタスクもしくはジョブに基づくものである。このため、パラメータの解析によって、ログエントリを、特定のインスタンスに属するものとして指定することができる。ログエントリとインスタンス指定との対応性を、様々な技法を用いて達成することができる。１つの実施形態において、ログエントリのセットについてベクトルが生成される。ベクトルは、セット内の各ログエントリを計上する。例えば、ログエントリ

をパースし、テンプレートｉｄにマッピングすることができる。次に、ベクトルは、このテンプレートに対応するカウンタをインクリメントさせる。このようにして、ベクトルを、テンプレートのバッグ（Ｂａｇ　ｏｆ　Ｔｅｍｐｌａｔｅｓ）として考えることができる。いくつかの実施形態では、ステップ２１０において、ログエントリのさらなるフィルタリングまたは強化を行うことができる。例えば、各次元のカウントは、ログまたは他の機能を用いて変換することができるか、またはログエントリは、テンプレートマッピング以外の方式でグループ化することができる。

【００２５】

ステップ２１５において、各インスタンスは、そのメッセージのテンプレートのセットによって表される。この結果、テキスト解析において用いられる「単語のバッグ（ｂａｇ－ｏｆ－ｗｏｒｄｓ）」表現に類似した「テンプレートのバッグ」表現が生じる。当該技術分野において一般的に理解されるように、単語のバッグは、センテンスまたは文書等のテキスト内で単語がどの程度頻繁に生じるかを表すモデルである。通常の実施態様において、テキストは固定長のベクトルに変換され、ここで、ベクトル内の各エントリは、単語の所定の辞書内の単語に対応する。フィールド内に記憶される数は、テキスト内で単語がどの程度頻繁に生じるかを表す。

【００２６】

ステップ２２０において、１つまたはそれ以上のクラスタリング解析がインスタンスに対し実行される。クラスタリングのために様々な技法を用いることができる。これらの技法は、直接、またはｓｃｉｋｉｔ－ｌｅａｒｎ等の標準的なツールキットを通じて実施することができる。図３は、ジャッカード類似度を用いたワンパスクラスタリングを実行する例示的な方法３００を示す。当該技術分野において一般的に理解されるように、セットのジャッカード類似度は、セットの和集合のサイズに対する積集合のサイズの比率である。ステップ３０５において、空のクラスタリストが作成される。次に、ステップ３１０において、第１のインスタンスが新たなクラスタのプロトタイプとして宣言され、このインスタンスがクラスタリストに加えられる。次に、他のインスタンスにおけるインスタンスｌについてステップ３１５～３３０が実行される。ステップ３１５から開始して、コンピューティングシステムは、ｌと既存のクラスタの各々とのジャッカード類似度を計算する。ステップ３２０において、最も高い類似度ｓを有するクラスタl_sが選択される。ステップ３２４において、「ｓ」の値が所定の閾値よりも大きい場合、ｌがクラスタl_sに加えられる。ステップ３３０において、ｓの値が閾値以下である場合、ｌについて新たなクラスタが作成され、このクラスタがクラスタリストに加えられる。

【００２７】

図２に戻ると、本明細書において「重要度値」と呼ばれるクラスタ確率値が、ステップ２２５において各テンプレートについて計算される。より詳細には、任意のテンプレートｋおよびクラスタｊについて、このクラスタ内でテンプレートを見る確率、およびこのテンプレートが任意の他のクラスタ内にある確率が計算される。これらの値の間の差異は、クラスタｊ内のテンプレートｋの重要度である。値が０に近づく場合、テンプレートはクラスタ内において外側と同じだけ一般的である。この値が高い場合、テンプレートは、クラスタ内において外側よりもはるかにより一般的であり、低い場合、外側においてはるかにより一般的である。重要度測定値は、クラスタがテンプレートの観点で表されることを可能にし、エンドユーザがクラスタリング結果を自身の領域知識と連結するのに役立つ。特に、大量のインスタンス、したがって最大クラスタが正常動作に対応する場合、いくつかのより小さなクラスタは異常インスタンスからなり、テンプレートの観点からのそれらの「シグネチャ」は、ユーザに異常の症状または原因を指し示す。

【００２８】

他の実施形態では、図２に示す方法２００とは別のログ解析の異なる方法を適用することができる。例えば、いくつかの実施形態では、ステップ２１０からの表現に外れ値／異常検出方法を適用することができ、その後に、結果を解釈するための説明可能な方法の使用が続く。

【００２９】
いくつかの実施形態では、視覚インタフェースを用いて、人間のユーザがログエントリについて生成されたクラスタリングおよび重要度値を理解するのを支援する。図４は、本明細書において論じられる解析のために用いることができる例示的なＧＵＩを示す。この例において、インスタンスは、プロット上の円形シンボルで示され、ここで、ｘ軸はインスタンスの時間を示し、ｙ軸はインスタンス識別子を示す。この場合、ログエントリは免疫測定法および臨床化学分析器システム（図１を参照）に対応し、インスタンス識別子は試料識別子である。キー４０５に示すように、各シンボルは、特定の試料がいずれのクラスタに属するかを示す。代替的に、シンボルは、例えば、異常検出アルゴリズムの出力を反映することができる。ユーザは、選択されたインスタンスに焦点を当てるために、いくつかのクラスタを起動する（停止させる）ことができる。例えば、図４において、ユーザは、０ｘ０４として指定され、図面におけるクラスタ３に属する、特定のインスタンスを選択した。インスタンスに関するさらなる詳細を与えるグラフィカル要素４１０が表示される。さらに、下部ディスプレイ４１５は、日付およびタイムスタンプ、対応するテンプレートの重要度値（すなわち、「重み」）、およびログメッセージ自体を含む、インスタンス内の各ログエントリに関する追加情報を提供する。このため、このＧＵＩは、ユーザが異常なインスタンスを、それらが最初に現れたときに迅速に識別することを可能にする。ユーザは、図５に示すもの等の特徴量重要度プロットをチェックすることによって、異常なインスタンスが、何により他のクラスタと異なっているかをチェックすることもできる。

【００３０】
図６は、本発明の実施形態を実施することができる例示的なコンピューティング環境６００を示す。例えば、いくつかの実施形態では、コンピューティング環境６００は、図１に示すログ処理コンピュータ１２５のために用いることができる。コンピューティング環境６００は、本発明の実施形態を実施することができるコンピューティングシステムの１つの例であるコンピュータシステム６１０を含むことができる。コンピュータシステム６１０およびコンピューティング環境６００等のコンピュータおよびコンピューティング環境は、当業者には既知であるため、ここでは簡潔に記載する。

【００３１】
図６に示すように、コンピュータシステム６１０は、コンピュータシステム６１０内で情報を通信するために、バス６２１または他の通信機構等の通信機構を含むことができる。コンピュータシステム６１０は、情報を処理するために、バス６２１と連結された１つまたはそれ以上のプロセッサ６２０をさらに含む。プロセッサ６２０は、１つもしくはそれ以上の中央処理装置（ＣＰＵ）、画像処理装置（ＧＰＵ）、または当該技術分野において既知の任意の他のプロセッサを含むことができる。

【００３２】
コンピュータシステム６１０は、バス６２１に連結され、情報およびプロセッサ６２０によって実行される命令を記憶する、システムメモリ６３０も含む。システムメモリ６３０は、リードオンリーメモリ（ＲＯＭ）６３１および／またはランダムアクセスメモリ（ＲＡＭ）６３２等の、揮発性および／または不揮発性メモリの形態のコンピュータ可読記憶媒体を含むことができる。システムメモリＲＡＭ６３２は、他のダイナミック記憶デバイス（例えば、ダイナミックＲＡＭ、スタティックＲＡＭ、およびシンクロナスＤＲＡＭ）を含むことができる。システムメモリＲＯＭ６３１は、他のスタティック記憶デバイス（例えば、プログラマブルＲＯＭ、消去可能ＰＲＯＭ、および電気的消去可能ＰＲＯＭ）を含むことができる。加えて、システムメモリ６３０を用いて、プロセッサ６２０による命令の実行中の一時変数または他の中間情報を記憶することができる。例えば、起動中に、コンピュータシステム６１０内の要素間で情報を転送するのに役立つ基本的なルーチンを含む基本的な入出力システム（ＢＩＯＳ）６３３をＲＯＭ６３１に記憶することができる。ＲＡＭ６３２は、プロセッサ６２０によって即座にアクセス可能および／または現在動作中のデータおよび／またはプログラムモジュールを含むことができる。システムメモ

リ６３０は、例えば、オペレーティングシステム６３４、アプリケーションプログラム６３５、他のプログラムモジュール６３６、およびプログラムデータ６３７をさらに含むことができる。

【００３３】

コンピュータシステム６１０は、ハードディスク６４１および取外し可能媒体ドライブ６４２（例えば、フロッピーディスクドライブ、コンパクトディスクドライブ、テープドライブ、および／またはソリッドステートドライブ）等の、情報および命令を記憶する１つまたはそれ以上の記憶デバイスを制御するために、バス６２１に連結されるディスク制御装置６４０も含む。適切なデバイスインタフェース（例えば、スモール・コンピュータ・システム・インタフェース（ＳＣＳＩ）、集積デバイス・エレクトロニクス（ＩＤＥ）、ユニバーサルシリアルバス（ＵＳＢ）、またはＦｉｒｅＷｉｒｅ）を用いて、記憶デバイスをコンピュータシステム６１０に追加することができる。

【００３４】

コンピュータシステム６１０はまた、コンピュータのユーザに対して情報を表示するための、陰極線管（ＣＲＴ）または液晶ディスプレイ（ＬＣＤ）等のディスプレイ６６６を制御するために、バス６２１に連結されたディスプレイコントローラ６６５を含むことができる。コンピュータシステムは、入力インタフェース６６０と、コンピュータのユーザとインタラクトして、プロセッサ６２０に情報を提供するための、キーボード６６２およびポインティングデバイス６６１等の１つまたは複数の入力装置とを含む。ポインティングデバイス６６１は、例えば、プロセッサ６２０に対して方向情報およびコマンド選択を伝達して、ディスプレイ６６６上のカーソルの動きを制御する、マウス、トラックボール、またはポインティングスティックとすることができる。ディスプレイ６６６は、ポインティングデバイス６６１による方向情報およびコマンド選択の伝達を補うまたはそれに取って代わる入力を可能にするタッチスクリーンインタフェースを提供することができる。

【００３５】

システムメモリ６３０等のメモリ内に含まれる１つまたはそれ以上の命令の１つまたはそれ以上のシーケンスをプロセッサ６２０が実行することに応答して、コンピュータシステム６１０は、本発明の実施形態の処理ステップの一部または全部を実行することができる。そのような命令は、ハードディスク６４１または取外し可能媒体ドライブ６４２等の別のコンピュータ可読媒体からシステムメモリ６３０内に読み込むことができる。ハードディスク６４１は、本発明の実施形態によって用いられる１つまたはそれ以上のデータストアおよびデータファイルを含むことができる。データストアコンテンツおよびデータファイルは、安全性向上のために暗号化することができる。プロセッサ６２０は、システムメモリ６３０に含まれる命令の１つまたはそれ以上のシーケンスを実行するために、多重処理構成で使用することができる。代替的な実施形態では、ソフトウェア命令の代わりにまたはソフトウェア命令と組み合わせて、ハードワイヤ有線回路を用いることができる。したがって、実施形態は、ハードウェア回路およびソフトウェアの任意の特定の組合せに限定されない。

【００３６】

上述のように、コンピュータシステム６１０は、本発明の実施形態に従ってプログラムされた命令を保持して、データ構造、テーブル、レコード、または本明細書に記載されている他のデータを含む、少なくとも１つのコンピュータ可読媒体またはメモリを含むことができる。本明細書において用いられるとき、「コンピュータ可読媒体」という用語は、プロセッサ６２０に対して命令を与えて、実行させることに関与する任意の媒体を指す。コンピュータ可読媒体は、限定ではないが、不揮発性媒体、揮発性媒体、および伝送媒体を含む多くの形態をとることができる。不揮発性媒体の非限定的な例として、ハードディスク６４１または取外し可能媒体ドライブ６４２等の、光ディスク、ソリッドステートドライブ、磁気ディスク、および光磁気ディスクが挙げられる。揮発性媒体の非限定的な例として、システムメモリ６３０等のダイナミックメモリが挙げられる。伝送媒体の非限定的な例として、バス６２１を構築するワイヤを含む、同軸ケーブル、銅線、および光ファ

イバーが挙げられる。伝送媒体は、電波データ通信中および赤外線データ通信中に生じるような音波または光波の形態をとることもできる。

【００３７】

コンピューティング環境６００は、リモートコンピュータ６８０等の１つまたはそれ以上のリモートコンピュータとの論理的接続を用いてネットワーク環境で動作するコンピュータシステム６１０をさらに含むことができる。リモートコンピュータ６８０は、パーソナルコンピュータ（ラップトップまたはデスクトップ）、モバイルデバイス、サーバ、ルータ、ネットワークＰＣ、ピアデバイスまたは他の一般的なネットワークノードとすることができ、典型的には、コンピュータシステム６１０に関して上述した要素の多くまたは全てを含む。ネットワーク環境内で用いられる場合、コンピュータシステム６１０は、インターネット等のネットワーク６７１を介した通信を確立するためのモデム６７２を含むことができる。モデム６７２は、ユーザネットワークインタフェース６７０を介して、または別の適切な機構を介してバス６２１に連結することができる。

【００３８】

ネットワーク６７１は、当技術分野で一般的に知られている任意のネットワークまたはシステムとすることができ、インターネット、イントラネット、ローカルエリアネットワーク（ＬＡＮ）、ワイドエリアネットワーク（ＷＡＮ）、メトロポリタンネットワーク（ＭＡＮ）、直接接続または一連の接続、携帯電話ネットワーク、またはコンピュータシステム６１０と他のコンピュータ（例えば、リモートコンピュータ６８０）との間の通信を促進することが可能な任意の他のネットワークもしくは媒体を含む。ネットワーク６７１は、有線、無線、またはそれらの組合せとすることができる。有線接続は、イーサネット、ユニバーサルシリアルバス（ＵＳＢ）、ＲＪ－１１または当技術分野で一般的に既知の任意の他の有線接続を用いて実施することができる。無線接続は、Ｗｉ－Ｆｉ、ＷｉＭＡＸ、およびブルートゥース、赤外線、携帯電話ネットワーク、通信衛星、または当技術分野で一般的に既知の任意の他の無線接続方法を用いて実施することができる。さらに、いくつかのネットワークが、単独または互いに通信して、ネットワーク６７１における通信を促進することができる。

【００３９】

本開示の実施形態は、ハードウェアおよびソフトウェアの任意の組合せによって実施することができる。さらに、本開示の実施形態は、例えば、コンピュータ可読の非一時的媒体を有する製造物（例えば、１つまたはそれ以上のコンピュータプログラム製品）に含めることができる。この媒体は、例えば、本開示の実施形態の機構を提供および促進するためのコンピュータ可読プログラムコードを内蔵している。この製造物は、コンピュータシステムの一部として含めることもできるし、または、別個に販売することもできる。

【００４０】

本明細書には様々な態様および実施形態が開示されているが、当業者には、他の態様および実施形態が明らかとなろう。本明細書に開示された様々な態様および実施形態は、例示を目的としており、限定を意図したものではない。真の範囲および趣旨は、添付の特許請求の範囲に示されている。

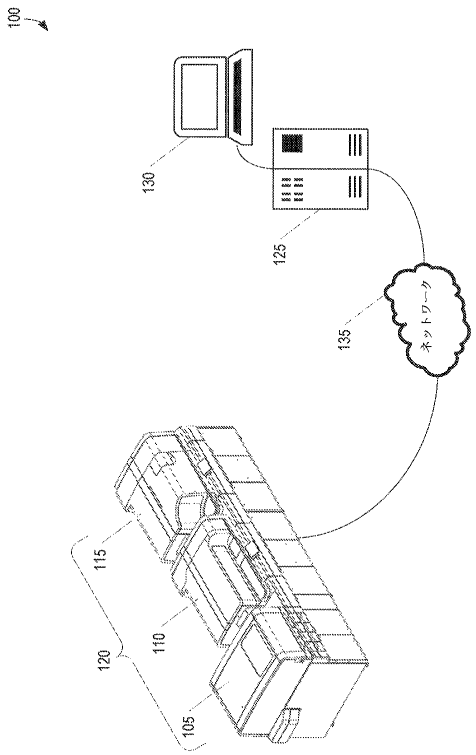
【００４１】

本明細書において用いられるとき、実行可能なアプリケーションは、オペレーティングシステム、コンテキストデータ収集システムまたは他の情報処理システムの機能等の、所定の機能を、例えばユーザコマンドまたは入力に応答して実施するように、プロセッサを条件付けるコードまたは機械可読命令を含む。実行可能手順は、コードまたは機械可読命令、サブルーチン、または１つもしくはそれ以上の特定のプロセスを実行するための実行可能なアプリケーションのコードまたは一部の他の別々のセクションのセグメントである。これらのプロセスは、入力データおよび／またはパラメータを受信すること、受信された入力データについて操作を実行すること、および／または受信された入力パラメータに応答して機能を実行すること、および結果として得られる出力データおよび／またはパラメータを提供することを含み得る。

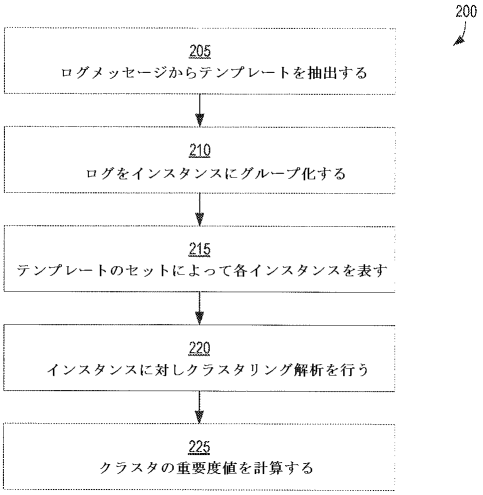
【００４２】
本明細書において用いられるとき、「グラフィカルユーザインタフェース（ＧＵＩ）」という用語は、ディスプレイプロセッサによって生成されて、プロセッサまたは他のデバイスとのユーザインタラクションならびに関連するデータ収集および処理機能を可能にする、１つまたは複数の表示画像を含む。ＧＵＩは、実行可能手順または実行可能アプリケーションも含む。実行可能手順または実行可能アプリケーションは、ＧＵＩ表示画像を表す信号を生成するために、ディスプレイプロセッサを条件付ける。これらの信号は、ユーザが見る画像を表示するディスプレイデバイスに供給される。プロセッサは、実行可能手順または実行可能アプリケーションの制御下で、入力デバイスから受信した信号に応答してＧＵＩ表示画像を操作する。このようにして、ユーザは、入力デバイスを用いて表示画像とインタラクトすることができ、プロセッサまたは他の装置とのユーザインタラクションが可能になる。
【００４３】
本明細書に記載の機能およびプロセスのステップは、自動的に、またはユーザコマンドに応答して全体的もしくは部分的に実施することができる。自動的に実施される行為（ステップを含む）は、ユーザがその行為を直接的に始動させなくても、１つまたはそれ以上の実行可能命令またはデバイス動作に応答して実行される。
【００４４】
図示したシステムおよびプロセスは排他的でない。本発明の原理に従って同一の目的を達成するため、他のシステム、プロセスおよびメニューを導出することができる。本発明は特定の実施形態を参照して説明されたが、本明細書で示し記載した実施形態および変形形態は例示を目的にしているにすぎないことを理解されたい。当業者であれば、現在の設計に対して、本発明の範囲から逸脱することなく、変更を行うことができる。本明細書に記載のように、様々なシステム、サブシステム、エージェント、マネージャーおよびプロセスは、ハードウェア構成要素、ソフトウェア構成要素、および／またはそれらの組合せを使用して実施することができる。請求項の要素はいずれも、「～するための手段」という語句を用いて明示的に記載されていない限り、米国特許法第１１２条（ｆ）の規定に基づいて解釈されるべきではない。

【図面】

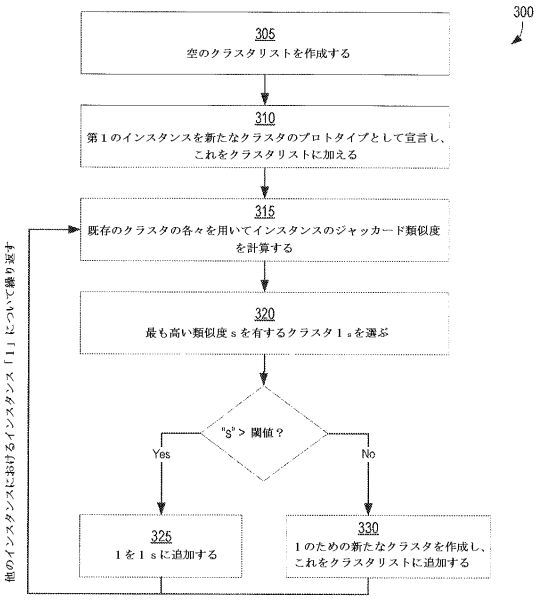
【図１】


100
130
125
135
ネットワーク
115
110
120
105

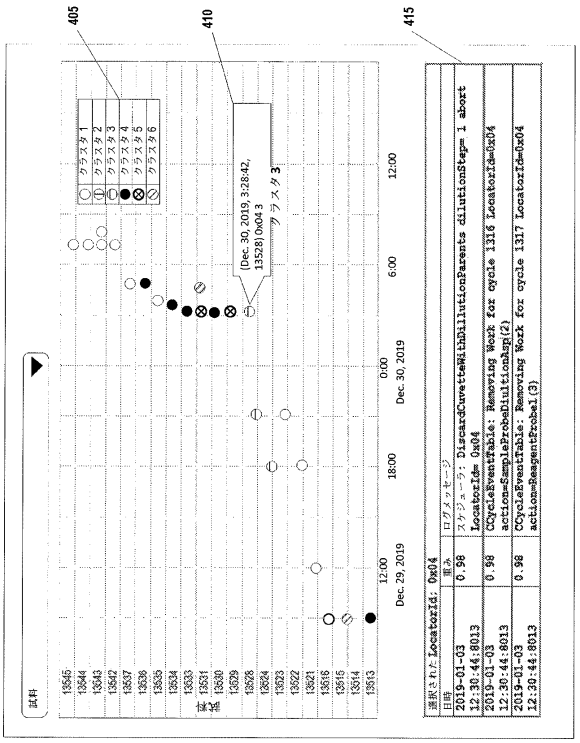
【図２】


200
205
ログメッセージからテンプレートを抽出する
210
ログをインスタンスにグループ化する
215
テンプレートのセットによって各インスタンスを表す
220
インスタンスに対しクラスタリング解析を行う
225
クラスタの重要度値を計算する

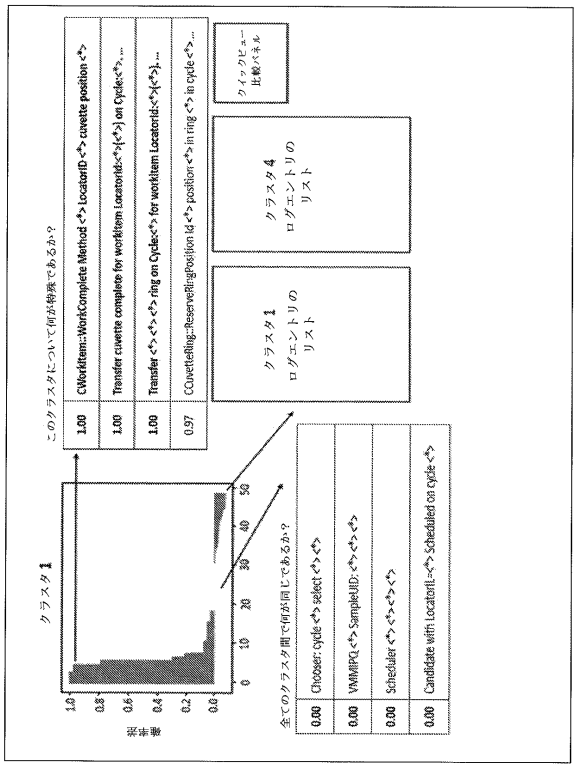
【図３】


300
305
空のクラスタリストを作成する
310
第１のインスタンスを新たなクラスタのプロトタイプとして宣言し、これをクラスタリストに加える
315
既存のクラスタの各々を用いてインスタンスのジャッカード類似度を計算する
320
最も高い類似度ｓを有するクラスタｌｓを選ぶ
"s" > 閾値？
Yes
No
325
ｌをｌｓに追加する
330
ｌのための新たなクラスタを作成し、これをクラスタリストに追加する
他のインスタンスにおけるインスタンス「ｉ」について繰り返す

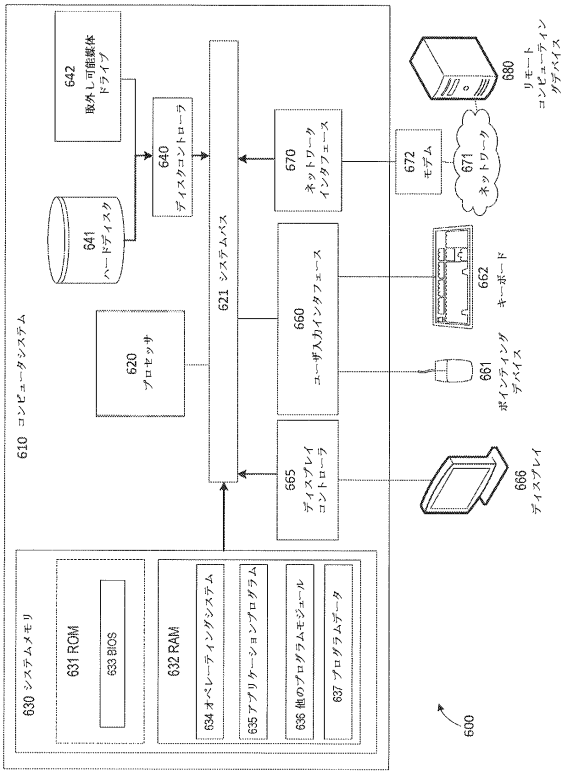
【図４】


405
410
415
クラスタ 1
クラスタ 2
クラスタ 3
クラスタ 4
クラスタ 5
クラスタ 6
(Dec. 30, 2019, 3:28:42, 13528) 0x004 3
クラスタ 3
12:00
6:00
0:00
Dec. 30, 2019
18:00
12:00
Dec. 29, 2019
試料
球塗
13546
13544
13543
13542
13537
13536
13535
13534
13533
13531
13530
13529
13528
13524
13523
13522
13521
13516
13515
13514
13513
選択された LocatorId: 0x04
日時
重み
ログメッセージ
2019-01-03 12:30:44:8013 0.98 セッション: DiscardCuvetteWithDilutionParents dilutionStage= 1 abort LocatorId= 0x04
2019-01-03 12:30:44:8013 0.98 CCycleEventTable: Removing Work for cycle 1316 LocatorId=0x04 action=SampleProbeDilutionMag(2)
2019-01-03 12:30:44:8013 0.98 CCycleEventTable: Removing Work for cycle 1317 LocatorId=0x04 actionReagentProbe1(3)

【図５】


クラスタ1
このクラスタについて何が特殊であるか？
1.00 CWorkItem::WorkComplete Method <*> LocationID <*> cuvette position <*>
1.00 Transfer cuvette complete for workItem LocationId:<*>{<*>} on Cycle:<*>, ...
1.00 Transfer <*> <*> <*> ring on Cycle:<*> for workItem LocationId:<*>{<*>} ...
0.97 CCuvetteRing::ReserveRingPosition Id <*> position <*> in ring <*> in cycle <*> ...
クイックビュー 比較パネル
クラスタ1 ログエントリのリスト
クラスタ4 ログエントリのリスト
全てのクラスタ間で何が同じであるか？
0.00 Chooser: cycle <*> select <*> <*>
0.00 VMMRPQ <*> SampleUID: <*> <*> <*>
0.00 Scheduler <*> <*> <*> <*> <*>
0.00 Candidate with LocationId:=<*> Scheduled on cycle <*>
重要度
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50

【図６】


610 コンピュータシステム
620 プロセッサ
621 システムバス
630 システムメモリ
631 ROM
633 BIOS
632 RAM
634 オペレーティングシステム
635 アプリケーションプログラム
636 他のプログラムモジュール
637 プログラムデータ
640 ディスクコントローラ
641 ハードディスク
642 取外し可能媒体ドライブ
660 ユーザ入力インタフェース
661 ポインティングデバイス
662 キーボード
665 ディスプレイコントローラ
666 ディスプレイ
670 ネットワークインタフェース
671 ネットワーク
672 モデム
680 リモートコンピューティングデバイス
600

フロントページの続き

５３６．プレインズボロ．フォックス・ラン・ドライブ２６１０
(72)発明者　オラディメジ・ファリ
アメリカ合衆国ニュージャージー州０７４５８．アッパー・サドル・リバー．ヒッコリー・プレイス７

審査官　渡辺 順哉

(56)参考文献
国際公開第２０１１／１１１５９９（WO，Ａ１）
特開２０１７－１６７８８０（ＪＰ，Ａ）
特開２０１７－０６８７４８（ＪＰ，Ａ）
特開２０１４－０３５７４９（ＪＰ，Ａ）
特表２０１８－５３０８０３（ＪＰ，Ａ）
特開２０１９－０５３７６３（ＪＰ，Ａ）
特表２０１９－５２３４０９（ＪＰ，Ａ）
米国特許出願公開第２０１１／０１８５２３４（ＵＳ，Ａ１）

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１１／０７－１１／３６
Ｇ０６Ｆ　１８／２３