(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
2016 年 12 月 15 日(15.12.2016)

WIPO | PCT

(10) 国際公開番号
WO 2016/199433 A1

(51) 国際特許分類:
G06F 11/34 (2006.01)

(21) 国際出願番号: PCT/JP2016/002816

(22) 国際出願日: 2016 年 6 月 10 日(10.06.2016)

(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 2015-118217 2015 年 6 月 11 日(11.06.2015) 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). ＮＥＣソリューションイノベータ株式会社(NEC SOLUTION INNOVATORS, LTD.) [JP/JP]; 〒1368627 東京都江東区新木場一丁目１８番７号 Tokyo (JP).

(72) 発明者: 網代 育大(AJIRO, Yasuhiro); 〒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). 鳥山 慎一(TORIYAMA, Shinichi); 〒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). 藤田 和也(FUJITA, Kazuya); 〒1368627 東京都江東区新木場一丁目１８番７号ＮＥＣソリューションイノベータ株式会社内 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, 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.

[続葉有]

(54) **Title**: MESSAGE ANALYSIS APPARATUS, MESSAGE ANALYSIS METHOD, AND STORAGE MEDIUM

(54) 発明の名称 : メッセージ分析装置、メッセージ分析方法、および、記憶媒体

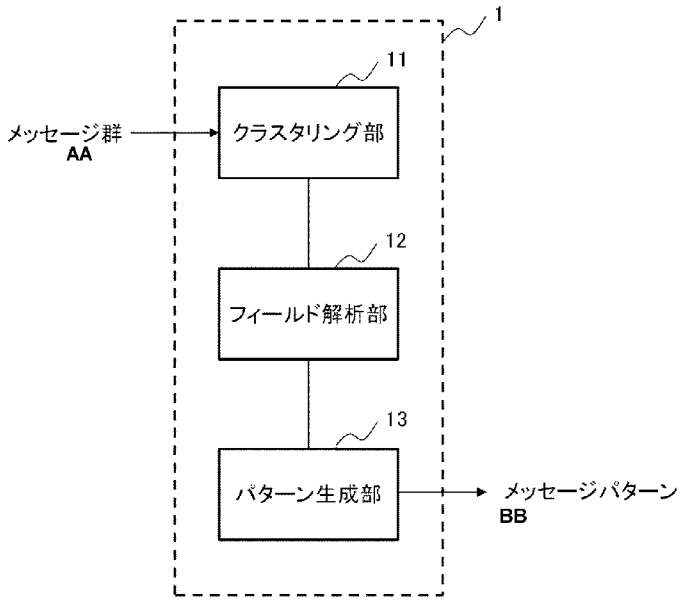


FIG. 1:
11 Clustering unit
12 Field analysis unit
13 Pattern generation unit
AA Message group
BB Message pattern

(57) **Abstract**: In the present invention, information indicating the contents or trends of many messages can be provided without the need of pre-defining the parts that vary among the messages. A message analysis apparatus 1 is provided with a clustering unit 11, a field analysis unit 12, and a pattern generation unit 13. The clustering unit 11 classifies a message group, which is a collection of messages comprising one or more fields, into clusters on the basis of similarities among the messages. The field analysis unit 12 identifies, for each of the fields constituting the message group in a cluster, a variable part in which the value thereof varies and an invariable part in which the value thereof does not vary. The pattern generation unit 13 generates a message pattern common to the messages in the cluster on the basis of the variable part and the invariable part.

(57) 要約: 本発明は、メッセージ間で変異する部分をあらかじめ定義する必要なく、多数のメッセージの内容や傾向を表す情報を提示することができる。 メッセージ分析装置１は、クラスタリング部１１と、フィールド解析部１２と、パターン生成部１３とを備える。クラスタリング部１１は、１つ以上のフィールドからなるメッセージの集合であるメッセージ群を、メッセージ間の類似性に基づいてクラスタに分類する。フィールド解析部１２は、クラスタ内のメッセージ群を構成する各フィールドについて、その値が変異している変数部分と、変異していない不変部分とを識別する。パターン生成部１３は、変数部分および不変部分に基づいて、クラスタ内のメッセージ群に共通するメッセージパターンを生成する。

WO 2016/199433 A1

(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:

— 国際調査報告（条約第 21 条(3)）

明　細　書

発明の名称 ：

メッセージ分析装置、メッセージ分析方法、および、記憶媒体

技術分野

[0001] 本発明は、多数のメッセージを分析する技術に関する。

背景技術

[0002] 一般に、装置やサービスでは、その動作状況や利用状況の履歴として、ログと呼ばれるメッセージが大量に記録される。また、インターネット上のソーシャルネットワークサービス等では、多数のユーザによりメッセージが入力され記録される。このような多数のメッセージを分析する分析者には、大量のメッセージの中に含まれる情報の内容や傾向を把握することが求められる。

[0003] メッセージを分析する技術の一例が、特許文献１に記載されている。特許文献１に記載された関連技術は、ログに含まれるメッセージから、他のメッセージと共通する共通部分と、他のメッセージと相違する相違部分とを抽出する。そして、この関連技術は、抽出した共通部分に識別情報を付与して共通部分情報として格納し、抽出した相違部分に識別情報を付与して相違部分情報として格納する。そして、この関連技術は、各メッセージを、共通部分の識別情報および相違部分の識別情報と関連付けて格納する。この関連技術を用いれば、メッセージの分析者は、大量のメッセージの中の共通部分および相違部分を把握することができる。

先行技術文献

特許文献

[0004] 特許文献1：国際公開第２０１３／１３６４１８号

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、特許文献１に記載された関連技術は、共通部分および相違

部分を抽出するために、相違部分を構成する変数の定義を必要とする。例えば、オペレーティングシステムの動作記録としてのログに含まれるメッセージに関して、プロセスＩＤを表す変数として、一文字以上の数字列が定義される。また、ＩＰ（Internet Protocol）アドレスを表す変数として、ピリオドで区切られた数字列が定義される。そして、この関連技術は、メッセージのうち変数の定義に一致する部分を相違部分として抽出し、その他の部分を共通部分として抽出する。このように、この関連技術は、あらかじめ変数を定義しておかなければ、大量のメッセージの共通部分および相違部分を抽出することができず、その内容や傾向を表す情報として提示できない。

[0006] 本発明は、上述の課題を解決するためになされたものである。すなわち、本発明は、メッセージ間で変異する部分をあらかじめ定義する必要なく、多数のメッセージの内容や傾向を表す情報を提示する技術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

[0007] 上記目的を達成するために、本発明のメッセージ分析装置は、１つ以上のフィールドからなるメッセージの集合であるメッセージ群を、前記メッセージ間の類似性に基づいてクラスタに分類するクラスタリング手段と、前記クラスタ内のメッセージ群を構成する各フィールドについて、その値が変異している変数部分と、変異していない不変部分とを識別するフィールド解析手段と、前記変数部分および前記不変部分に基づいて、前記クラスタ内のメッセージ群に共通するメッセージパターンを生成するパターン生成手段と、を備える。

[0008] また、本発明のメッセージ分析方法は、コンピュータ装置を用いて、１つ以上のフィールドからなるメッセージの集合であるメッセージ群を、前記メッセージ間の類似性に基づいてクラスタに分類し、前記クラスタ内のメッセージ群を構成する各フィールドについて、その値が変異している変数部分と、変異していない不変部分とを識別し、前記変数部分および前記不変部分に基づいて、前記クラスタ内のメッセージ群に共通するメッセージパターンを

生成する。

[0009] また、本発明の記憶媒体は、１つ以上のフィールドからなるメッセージの集合であるメッセージ群を、前記メッセージ間の類似性に基づいてクラスタに分類するクラスタリングステップと、前記クラスタ内のメッセージ群を構成する各フィールドについて、その値が変異している変数部分と、変異していない不変部分とを識別するフィールド解析ステップと、前記変数部分および前記不変部分に基づいて、前記クラスタ内のメッセージ群に共通するメッセージパターンを生成するパターン生成ステップと、をコンピュータ装置に実行させるメッセージ分析プログラムを記憶している。

発明の効果

[0010] 本発明は、メッセージ間で変異する部分をあらかじめ定義する必要なく、多数のメッセージの内容や傾向を表す情報を提示する技術を提供することができる。

図面の簡単な説明

[0011] [図1]本発明の第１の実施の形態としてのメッセージ分析装置の構成を示すブロック図である。

[図2]本発明の第１の実施の形態としてのメッセージ分析装置のハードウェア構成の一例を示す図である。

[図3]本発明の第１の実施の形態としてのメッセージ分析装置の動作を説明するフローチャートである。

[図4]本発明の第２の実施の形態としてのメッセージ分析装置の構成を示すブロック図である。

[図5]本発明の第２の実施の形態としてのメッセージ分析装置の動作を説明するフローチャートである。

[図6]本発明の第２の実施の形態におけるクラスタリング結果の具体例を示す図である。

[図7]本発明の第２の実施の形態におけるフィールド解析結果の具体例を示す図である。

[図8]本発明の第３の実施の形態としてのメッセージ分析装置の構成を示すブロック図である。

[図9]本発明の第３の実施の形態としてのメッセージ分析装置の動作を説明するフローチャートである。

[図10]本発明の第３の実施の形態において細分化されたクラスタの具体例を示す図である。

[図11]本発明の第４の実施の形態としてのメッセージ分析装置の構成を示すブロック図である。

[図12]本発明の第４の実施の形態としてのメッセージ分析装置の動作を説明するフローチャートである。

[図13]本発明の第４の実施の形態におけるフィールド解析結果の具体例を示す図である。

[図14]本発明の第４の実施の形態におけるフィールド間の相関の有無を模式的に説明する図である。

[図15]本発明の第４の実施の形態において細分化されたクラスタの具体例を示す図である。

発明を実施するための形態

[0012] 以下、本発明の実施の形態について、図面を参照して詳細に説明する。

[0013] （第１の実施の形態）

本発明の第１の実施の形態としてのメッセージ分析装置１の機能ブロック構成を図１に示す。図１において、メッセージ分析装置１は、クラスタリング部１１と、フィールド解析部１２と、パターン生成部１３とを備える。メッセージ分析装置１は、メッセージ群を分析してその内容や傾向を示すメッセージパターンを生成する装置である。

[0014] ここで、メッセージとは、装置やサービスまたは人間等によって記録される情報の単位をいう。例えば、メッセージは、装置やサービス等の動作状況や利用状況の履歴を表すログデータに含まれる情報の単位であってもよい。この場合、メッセージは、サーバやクライアントといったＩＴ（Information

Technology）システムの構成要素によって、所定のタイミング毎に生成され、ログデータに追加される単位の情報であってもよい。この場合、メッセージは、そのメッセージが出力された時刻や出力元の名称等を含んでいることが多い。また、この場合、メッセージは、ログデータを表すファイルに含まれる１行分のテキストデータであることが多い。ただし、１つのメッセージは、複数行にわたっていてもよい。あるいは、複数のメッセージが、１行に含まれていてもよい。例えば、ログデータを表すファイルにおいて、複数行に渡る１つのメッセージに含まれる改行コードを空白文字に変換する前処理や、１行に含まれる複数のメッセージ間の空白文字を改行コードに変換する前処理等が行われることを想定してもよい。この場合、メッセージは、ログデータを表すファイルの１行によって構成されるとみなすことができる。

[0015] その他、メッセージは、ログデータに含まれる情報に限らず、任意のサービスに対して入力装置やネットワークを介して入力され、記録される情報の単位であってもよい。

[0016] また、メッセージは、１つ以上のフィールドによって構成される。フィールドは、例えば、セパレータによって区切られた情報であってもよい。例えば、「April 1 13:31:52 logging start」というメッセージは、空白をセパレータとして、５つのフィールド「April」、「1」、「13:31:52」、「logging」、「start」から構成される。あるいは、日本語で構成されるメッセージのように、空白等のセパレータで区切られていないメッセージがある。このようなメッセージは、単語や形態素、カタカナやひらがな、漢字といった文字種で分離する前処理により、１つ以上のフィールドによって構成されるとみなすことができる。

[0017] つまり、本実施の形態におけるメッセージが、１つ以上のフィールドから構成されるという前提は、本実施の形態において処理可能なメッセージの種類を限定するものではない。どのような種類のメッセージであっても、必要に応じて前処理を施すことにより、１つ以上のフィールドから構成されるものとして処理可能である。

[0018] また、メッセージに対する前処理として、１つのフィールドを複数のフィールドに分割する処理も考えられる。例えば、あるフィールドの値が、あるメッセージでは「abc&def」であり、他のメッセージでは「abc&ghi」であったとする。また、メッセージの内容について、abc、defおよびghiが個別の対象を表すことが定義されていたとする。このような場合、「abc&def」は、１つのフィールドではなく、「abc」、「＆」および「def」のように３つのフィールドとして処理されるのに適している。メッセージに対する前処理には、このような処理が含まれ得る。

[0019] 本実施の形態では、必要に応じて上述した前処理が施された１つ以上のフィールドからなるメッセージの集合（対象のメッセージ群）が、メッセージ分析装置１に対して入力されるものとする。例えば、対象のメッセージ群は、各メッセージについてそのフィールドの値（文字列、数値、記号等）をテーブル形式で表した情報として、あらかじめ記憶装置に格納されていてもよい。

[0020] 次に、メッセージ分析装置１のハードウェア構成の一例を図２に示す。図２において、メッセージ分析装置１は、ＣＰＵ（Central Processing Unit）１００１、メモリ１００２、出力装置１００３、および、入力装置１００４を含む。メモリ１００２は、ＲＡＭ（Random Access Memory）、ＲＯＭ（Read Only Memory）、補助記憶装置（ハードディスク等）等によって構成される。出力装置１００３は、ディスプレイ装置やプリンタ等のように、情報を出力する装置によって構成される。入力装置１００４は、キーボードやマウス等のように、ユーザ操作の入力を受け付ける装置によって構成される。この場合、メッセージ分析装置１の各機能ブロックは、メモリ１００２に格納されるコンピュータ・プログラムを読み込んで実行するとともに出力装置１００３、入力装置１００４の各部を制御するＣＰＵ１００１によって構成される。なお、メッセージ分析装置１およびその各機能ブロックのハードウェア構成は、上述の構成に限定されない。

[0021] 次に、メッセージ分析装置１の各機能ブロックの詳細について説明する。

[0022] クラスタリング部１１は、対象のメッセージ群を、メッセージ間の類似性に基づいてクラスタに分類する。クラスタ数は、メッセージ数以下となる。なお、対象のメッセージ群は、前述のように、必要に応じて前処理を施された、１つ以上のフィールドからなるメッセージの集合である。例えば、クラスタリング部１１は、メモリ１００２にあらかじめ記憶された対象のメッセージ群を取得して、クラスタに分類してもよい。複数の情報をそれらの間の類似性に基づいて分類する手法には、公知の技術を採用可能である。

[0023] フィールド解析部１２は、クラスタ内のメッセージ群を構成する各フィールドについて、その値が変異している変数部分と、変異していない不変部分とを識別する。具体的には、フィールド解析部１２は、クラスタ内の全てのメッセージに渡って、その値が一致しているフィールドを不変部分として識別してもよい。また、フィールド解析部１２は、クラスタ内の全てのメッセージのうち、少なくともいずれかで値が異なるフィールドを変数部分として識別してもよい。

[0024] パターン生成部１３は、フィールドの変数部分と不変部分とに基づいて、クラスタ内のメッセージ群に共通するメッセージパターンを生成する。例えば、パターン生成部１３は、変数部分のフィールドを所定の記号（例えば、アスタリスク「＊」）で表した情報と、不変部分のフィールドをその値で表した情報とを、フィールドの出現順に並べた情報を、共通パターンとして生成してもよい。そして、パターン生成部１３は、そのクラスタに含まれるメッセージ群において、変数部分のフィールドがとる値のリストを抽出する。以降、変数部分として識別されたフィールドを変数と呼び、変数がとり得る値を引数と呼ぶことにする。そして、パターン生成部１３は、各クラスタについて、共通パターンおよび各変数の引数のリストを、メッセージパターンとして生成すればよい。

[0025] 以上のように構成されたメッセージ分析装置１の動作について、図３を参照して説明する。

[0026] まず、クラスタリング部１１は、対象のメッセージ群を、メッセージ間の

類似性に基づいてクラスタに分類する（ステップＳ１）。

[0027] 次に、フィールド解析部１２は、ステップＳ１で生成された各クラスタについて、クラスタ内のメッセージ群を構成する各フィールドのうち、その値が変異している変数部分と、変異していない不変部分とを識別する（ステップＳ２）。

[0028] 次に、パターン生成部１３は、各クラスタについて、変数部分と不変部分とに基づいて、クラスタ内のメッセージ群に共通するメッセージパターンを生成する（ステップＳ３）。

[0029] 前述のように、パターン生成部１３は、メッセージパターンとして、共通パターンおよび変数の引数のリストを生成してもよい。

[0030] 以上で、メッセージ分析装置１は、動作を終了する。

[0031] 次に、本発明の第１の実施の形態の効果について述べる。

[0032] 本発明の第１の実施の形態としてのメッセージ分析装置は、メッセージ間で変異する部分をあらかじめ定義する必要なく、多数のメッセージの内容や傾向を表す情報を提示することができる。

[0033] その理由について述べる。本実施の形態では、クラスタリング部が、メッセージ群を、メッセージ間の類似性に基づいてクラスタに分類する。そして、フィールド解析部が、クラスタ内のメッセージ群を構成する各フィールドについて、その値が変異している変数部分と、変異していない不変部分とを識別する。そして、パターン生成部が、フィールドの変数部分と不変部分とに基づいて、クラスタ内のメッセージ群に共通するメッセージパターンを生成するからである。

[0034] このように、本実施の形態は、メッセージ群において変異する部分の定義を必要とせずに、変異部分および不変部分を抽出可能である。これにより、本実施の形態は、事前の変数定義がなくても、類似するメッセージ群を、それらの間で不変の部分と変異する部分とを認識可能に利用者に提示することができる。その結果、本実施の形態を利用する利用者は、大量のメッセージ群の内容や傾向をより容易に把握することができる。

[0035] （第２の実施の形態）

次に、本発明の第２の実施の形態について図面を参照して詳細に説明する。なお、本実施の形態の説明において参照する各図面において、本発明の第１の実施の形態と同一の構成および同様に動作するステップには同一の符号を付して本実施の形態における詳細な説明を省略する。

[0036] まず、本発明の第２の実施の形態としてのメッセージ分析装置２の機能ブロック構成を図４に示す。図４において、メッセージ分析装置２は、本発明の第１の実施の形態としてのメッセージ分析装置１に対して、次の点が異なる。すなわち、メッセージ分析装置２は、クラスタリング部１１に替えてクラスタリング部２１と、フィールド解析部１２に替えてフィールド解析部２２と、パターン生成部１３に替えてパターン生成部２３とを備え、さらに、クラスタ類似性判定部２４を備える。なお、メッセージ分析装置２およびその各機能ブロックは、図２を参照して説明した本発明の第１の実施の形態と同一のハードウェア要素によって構成可能である。ただし、メッセージ分析装置２および各機能ブロックのハードウェア構成は、上述の構成に限定されない。

[0037] 次に、メッセージ分析装置２の各機能ブロックの詳細について説明する。

[0038] クラスタリング部２１は、あるメッセージと、そのメッセージに対して類似性が所定条件を満たす他のメッセージとを、同一クラスタに分類する。

[0039] 例えば、クラスタリング部２１は、２つのメッセージ間の類似性として、各メッセージを構成するフィールド数に対して一致するフィールド数の割合に基づく値（類似度）を用いてもよい。この場合、類似度が高いほど、２つのメッセージ間の類似性は高くなる。例えば、２つのメッセージが共に１０個のフィールドで構成され、そのうちの７個が一致する場合、これらのメッセージ間の類似度は、７／１０＝０．７と算出される。この場合、クラスタリング部２１は、あるメッセージと、そのメッセージに対して類似度が閾値以上の他の各メッセージとを、同一クラスタに分類してもよい。

[0040] あるいは、クラスタリング部２１は、２つのメッセージ間の類似性として

、各メッセージを構成するフィールド数に対して一致しないフィールド数の割合に基づく値（距離）を用いてもよい。この場合、距離が大きいほど、２つのメッセージ間の類似性は低くなる。例えば、２つのメッセージが共に１０個のフィールドで構成され、そのうちの３個が不一致の場合、これらのメッセージ間の距離は、３／１０＝０．３と算出される。この場合、クラスタリング部２１は、あるメッセージと、そのメッセージに対して距離が閾値以下の他の各メッセージとを、同一クラスタに分類してもよい。

[0041] なお、２つのメッセージのフィールド数が異なる場合、類似度または距離を算出する際の分母としては、多い方または少ない方のいずれかのフィールド数を採用することを、あらかじめ定めておけばよい。例えば、多い方のフィールド数を採用すると定められていることを想定する。このとき、９個のフィールドからなるメッセージと、１０個のフィールドからなるメッセージとの間で、６個のフィールドが等しいとする。この場合、これらのメッセージ間の類似度は、前述の計算手法の場合、６／１０＝０．６０と算出される。また、これらのメッセージ間の距離は、前述の計算手法の場合、４／１０＝０．４０と算出される。

[0042] また、クラスタリング部２１は、各メッセージにおいて所定のフィールドパターンに合致する部分を、メッセージ間で互いに類似するフィールドとみなして、メッセージ群をクラスタに分類する。ここで、所定のフィールドパターンとは、メッセージにおいて、値が異なっていても類似するフィールドであるとみなすことが可能な部分がとり得る値のパターンである。そのようなフィールドパターンは、あらかじめ定義されていてもよい。例えば、日付や日時等は、値が異なっていても類似するフィールドであるとみなすことができる。そこで、クラスタリング部２１は、日付形式や日時形式に合致するフィールドパターンをあらかじめ記憶しておいてもよい。そして、クラスタリング部２１は、上述の類似度や距離を算出する際に、これらのフィールドパターンに合致する部分については、値が異なっていても一致するフィールドであるとして算出を行えばよい。

[0043] クラスタ類似性判定部２４は、各クラスタについて、そのクラスタにおけるメッセージ群全体の類似性が所定条件を満たすか否かを判定する。以降、クラスタにおけるメッセージ群全体の類似性を、単に全体の類似性とも記載する。例えば、クラスタ類似性判定部２４は、クラスタ内におけるメッセージ群を構成するフィールドのうち、不変部分を構成するフィールドの割合を、全体の類似性として用いてもよい。この場合、所定条件とは、全体の類似性を表す値が閾値以上であることであってもよい。また、全体の類似性を表す値の閾値は、クラスタリング部２１によって２つのメッセージ間の類似性の判断に用いられた類似度の閾値と同一の値であってもよい。

[0044] 具体的には、クラスタ類似性判定部２４は、クラスタ内における不変部分を構成するフィールドの個数を、クラスタ内のメッセージのうちの最大フィールド数で除した値を、全体の類似性を表す値として算出してもよい。そして、この場合、クラスタ類似性判定部２４は、全体の類似性を表す値が閾値以上であるか否かを判定する。

[0045] ここで、クラスタリング部２１によって類似度または距離の閾値に基づき生成されたクラスタであっても、全体の類似性が所定条件を満たさない場合がある。これは、分類の際に基準としたメッセージに対して類似性があると判定された他の各メッセージによって、変異するフィールドが大きく異なる場合があるためである。そのようなクラスタは、メッセージパターンを生成する分類として適切ではないことが多い。そこで、クラスタ類似性判定部２４は、メッセージパターンの生成対象として適切でないクラスタを除外するために備えられた機能ブロックである。

[0046] なお、クラスタ類似性判定部２４によって全体の類似性が所定条件を満たさないと判定されたクラスタがあっても、全体の類似性が所定条件を満たすと判定された他のクラスタを対象として、後述のパターン生成部２３が処理を行ってもよい。あるいは、クラスタ類似性判定部２４によって全体の類似性が所定条件を満たさないと判定されたクラスタがある場合、クラスタリング部２１が、類似度の閾値を変更して再度クラスタリング処理をやり直して

もよい。

[0047] その場合、閾値の変更方法としては、閾値を上げる（大きくする）方法と、下げる（小さくする）方法がある。例えば、類似度に関する閾値を上げると、実際に出力されているメッセージ数に近い多数の細かいクラスタが得られる。つまり、最終的に得られるメッセージパターン数は、メッセージ数に対してより近くなる。また、類似度に関する閾値を下げると、実際に出力されているメッセージ数より少数の粗いクラスタが得られる。つまり、最終的に得られるメッセージパターン数は、メッセージ数に比べてより少なくなる。閾値の変更方法は、メッセージパターンの用途、メッセージ量、メッセージパターンの種類数等に応じて決定されてもよい。

[0048] パターン生成部２３は、クラスタ類似性判定部２４によって全体の類似性が所定条件を満たすと判定されたクラスタについて、本発明の第１の実施の形態におけるパターン生成部１３と同様にメッセージパターンを生成する。

[0049] 以上のように構成されたメッセージ分析装置２の動作について、図５を参照して説明する。

[0050] まず、クラスタリング部２１は、メッセージ群をクラスタリングする際の閾値を取得する（ステップＳ２１）。例えば、クラスタリング部２１は、入力装置１００４を介して閾値を取得してもよい。

[0051] 次に、クラスタリング部２１は、対象のメッセージ群のうち、あるメッセージと、そのメッセージに対して類似度が閾値以上または距離が閾値以下の他の各メッセージとを、同一のクラスタに分類する（ステップＳ２２）。

[0052] 具体的には、前述のように、クラスタリング部２１は、メッセージの集合からメッセージを１つ取り出し、このメッセージと、他の各メッセージとの間でそれぞれ類似度（または距離）を計算する。そして、クラスタリング部２１は、取り出したメッセージと、そのメッセージに対して閾値以上の類似度（または閾値以下の距離）が算出された各メッセージとによって、１つのクラスタを形成すればよい。

[0053] そして、クラスタリング部２１は、１つのクラスタを算出したのち、まだ

分類していない残りのメッセージについて同様の処理を行い、別のクラスタを形成する。そして、メッセージ分析装置２は、各クラスタについて、ステップＳ２３～Ｓ２７の処理を行う。

[0054] なお、メッセージ分析装置２は、まず全てのメッセージをいずれかのクラスタに分類してから、各クラスタについてステップＳ２３～Ｓ２７の処理を繰り返してもよい。あるいは、メッセージ分析装置２は、１つのクラスタを形成する度にそのクラスタについてステップＳ２３～Ｓ２７の処理を実行することを繰り返してもよい。

[0055] ここでは、まず、フィールド解析部２２は、クラスタ内の全てのメッセージに渡って値が一致しているフィールドと、フィールドパターンに一致するフィールドとを、不変部分として識別する。また、フィールド解析部２２は、少なくとも１つのメッセージにおいて値が異なるフィールドを、変数部分として識別する（ステップＳ２３）。

[0056] 次に、クラスタ類似性判定部２４は、このクラスタにおける全体の類似性が、所定条件を満たすか否かを判断する（ステップＳ２４）。

[0057] 前述のように、クラスタ類似性判定部２４は、このクラスタ中の不変部分を構成するフィールドの個数を最大フィールド数で除した値を、クラスタにおける全体の類似性を表す値として算出してもよい。そして、クラスタ類似性判定部２４は、このクラスタにおける全体の類似性を表す値が、閾値を超えるか否かを判断すればよい。

[0058] もし、このクラスタ全体の類似性が所定条件を満たさないと判断された場合、メッセージ分析装置２は、当該クラスタに関してメッセージパターンの生成に失敗した旨を出力して、処理を終了する。

[0059] 一方、このクラスタ全体の類似性が所定条件を満たすと判断された場合、パターン生成部２３は、このクラスタの共通パターンを生成する（ステップＳ２５）。

[0060] 具体的には、パターン生成部２３は、変数部分のフィールドを所定の記号（例えばアスタリスク「＊」）で表した情報と、不変部分のフィールドを表

す情報とを、フィールドの出現順に並べた情報を共通パターンとして生成する。なお、パターン生成部２３は、不変部分のうちフィールドパターンに一致するフィールドについては、その値の代わりに、所定の文字列を用いて共通パターンを生成してもよい。例えば、パターン生成部２３は、日付のフィールドパターンに一致するフィールドを「（Ｄａｔｅ）」と表し、時刻のフィールドパターンに一致するフィールドを「（Ｔｉｍｅ）」と表して、共通パターンを生成してもよい。

[0061] 次に、パターン生成部２３は、共通パターンのうち変数部分のフィールドについて、引数のリストを生成する（ステップＳ２６）。

[0062] そして、パターン生成部２３は、共通パターンおよび各変数部分の引数のリストを、このクラスタのメッセージパターンとして出力する（ステップＳ２７）。なお、出力先は、出力装置１００３、メモリ１００２、または、ネットワークを介して接続された他の装置等であってもよい。

[0063] 以上で、メッセージ分析装置２は、動作を終了する。

[0064] 次に、メッセージ分析装置２の動作を具体例で示す。

[0065] この具体例では、メッセージ分析装置２は、メッセージ間の類似性の判断に前述の類似度を用いるものとする。

[0066] ここでは、まず、クラスタリング部２１は、類似度の閾値として０．６を取得する（ステップＳ２１）。

[0067] 次に、クラスタリング部２１は、対象のログメッセージ群のうち、あるメッセージと、そのメッセージに対する他の各メッセージとの類似度を算出することにより、図６に示すクラスタＡおよびクラスタＢを形成する（ステップＳ２２）。

[0068] 図６において、各行は、１つのメッセージを表す。また、点線の楕円は、フィールドを示している。この例では、日付および時刻を表すフィールドパターンが定められている。クラスタリング部２１は、日付のフィールドパターンに一致している部分を日付フィールドとして、メッセージ間において互いに一致するものとみなす。また、クラスタリング部２１は、時刻のフィー

ルドパターンに一致している部分を時刻フィールドとして、メッセージ間において互いに一致するものとみなす。この場合、クラスタＡでは、１行目のメッセージに対して、２行目のメッセージは、９つのフィールドのうち７つが一致している。したがって、クラスタリング部２１は、１、２行目のメッセージ間の類似度として、７／９≒０．７７を算出する。このように、クラスタリング部２１は、１行目のメッセージと、１行目のメッセージに対して類似度が０．６以上の２～４行目の各メッセージとを、クラスタＡとして分類している。クラスタＢについても同様である。

[0069] 次に、メッセージ分析装置２は、クラスタＡについて、ステップＳ２３～Ｓ２７の処理を行う。

[0070] ここでは、フィールド解析部２２は、クラスタＡについて、不変部分のフィールドと、変数部分のフィールドとを識別し、図７に示す識別処理結果を生成する（ステップＳ２３）。

[0071] 図７において、まず、フィールド解析部２２は、識別処理結果を格納する表を作成する。識別処理結果を格納する表は、１列目（最も左の列）に、フィールドに付与したＩＤを含む。また、この表は、１行目（最も上の行）に、メッセージの識別情報を含む。また、この表は、２列目以降の各列に、各メッセージの解析結果を格納可能となっている。

[0072] 次に、フィールド解析部２２は、クラスタＡに含まれるメッセージの１つ（一例としてＭｓｇ１１３４）を代表メッセージとして識別処理を行う。まず、フィールド解析部２２は、代表メッセージＭｓｇ１１３４を構成する各フィールドの値を、図７の表の２列目に格納する。ただし、フィールド解析部２２は、日付のフィールドパターンに一致する日付フィールドについては、値の代わりに、日付を表す情報「（Ｄａｔｅ）」を格納する。また、フィールド解析部２２は、時刻のフィールドパターンに一致する時刻フィールドについては、値の代わりに、時刻を表す情報「（Ｔｉｍｅ）」を格納する。

[0073] 次に、フィールド解析部２２は、クラスタＡに含まれる次のメッセージＭｓｇ１２１１を構成する各フィールドの値のうち、代表メッセージと値が異

なるフィールドについて、３列目にその値を格納する。ただし、フィールド解析部２２は、日付フィールドおよび時刻フィールドについては、代表メッセージと値が一致するものとして、値を格納しない。そして、フィールド解析部２２は、クラスタＡの残りのメッセージＭｓｇ２０９１およびＭｓｇ４６２５についても同様に、代表メッセージと値が異なるフィールドについてその値を４列目および５列目に格納する。このようにして、フィールド解析部２２は、クラスタＡの全てのメッセージについて表に値を格納する処理を行い、図７の表を生成する。

[0074] 次に、フィールド解析部２２は、図７の表の３列目以降において少なくとも１つの列に値が格納されている４個のフィールド（フィールドＩＤ３、７、９、１０）を、クラスタＡの変数部分として識別する。また、フィールド解析部２２は、図７の表の３列目以降において値が格納されていない６個のフィールド（フィールドＩＤ１、２、４、５、６、８）を、クラスタＡの不変部分として識別する。

[0075] 次に、クラスタ類似性判定部２４は、クラスタＡにおける全体の類似性を表す値が、閾値以上であるか否かを判断する（ステップＳ２４）。

[0076] 図７を参照すると、クラスタＡでは、Ｍｓｇ２０９１を構成するフィールド数が最大で１０個である。また、ステップＳ２３において、６個のフィールド（フィールドＩＤ１、２、４、５、６、８）が、クラスタＡの不変部分として識別されている。したがって、クラスタ類似性判定部２４は、クラスタＡにおける全体の類似性を表す値として、６／１０＝０．６０を算出する。ここで、閾値が０．６であるため、クラスタ類似性判定部２４は、クラスタＡにおける全体の類似性を表す値は、閾値以上であると判断する。

[0077] そこで、パターン生成部２３は、クラスタＡの共通パターンを生成するため、不変部分のフィールドＩＤ１、２、４、５、６、８をその値またはフィールドパターンを表す情報で表現する。また、パターン生成部２３は、変数部分のフィールドＩＤ３、７、９、１０を所定の記号「＊」で表す。そして、パターン生成部２３は、これらの情報をフィールドＩＤの順に並べて、ク

ラスタＡの共通パターン「（Ｄａｔｅ）　（Ｔｉｍｅ）　＊　ｐｒｏｃｅｓｓ　ａｂｃ　［　＊　］　＊　＊」を生成する（ステップＳ２５）。

[0078] 次に、パターン生成部２３は、クラスタＡの共通パターンのうち変数部分のフィールドＩＤ３，７，９，１０について、それぞれ引数のリストを生成する（ステップＳ２６）。

[0079] 例えば、パターン生成部２３は、図７の表におけるフィールドＩＤ３の行を参照することにより、フィールドＩＤ３の引数のリスト「ｈｏｓｔ０１、ｈｏｓｔ０２、ｈｏｓｔ０３」を生成する。同様に、パターン生成部２３は、図７の表におけるフィールドＩＤ７、９、１０の行をそれぞれ参照することにより、引数のリストを生成する。

[0080] そして、パターン生成部２３は、クラスタＡの共通パターンおよび各変数部分の引数のリストを、メッセージパターンとして出力する（ステップＳ２７）。

[0081] そして、メッセージ分析装置２は、クラスタＢについてもステップＳ２３～Ｓ２７を実行する。

[0082] 以上で、具体例の説明を終了する。

[0083] 次に、本発明の第２の実施の形態の効果について述べる。

[0084] 本発明の第２の実施の形態としてのメッセージ分析装置は、大量のメッセージを、より少数のメッセージパターンの集合として提示することができ、メッセージの内容や傾向を利用者に対してより素早く把握させるよう支援することができる。

[0085] その理由について述べる。本実施の形態では、クラスタリング部が、メッセージ中の所定のフィールドパターンに一致する部分については類似するフィールドとみなしてクラスタリングを行うからである。また、フィールド解析部が、所定のフィールドパターンに一致する部分については不変部分とみなして共通メッセージを生成するからである。

[0086] これにより、本実施の形態は、複数のメッセージ間における軽微な相違については類似するものとみなすことができ、軽微な相違も変数部分とみなす

場合と比べて、より少数の共通するメッセージパターンを生成することができる。

[0087] また、他の理由について述べる。本実施の形態では、クラスタ類似性判定部が、クラスタにおける全体の類似性が所定条件を満たすか否かを判断するからである。そして、パターン生成部が、全体の類似性が所定条件を満たすクラスタについて、メッセージパターンを生成するからである。

[0088] これにより、本実施の形態は、全体の類似性が適正なクラスタについてメッセージパターンを生成するので、メッセージ群の内容と傾向をより精度よく反映したメッセージパターンを提示することができる。

[0089] （第３の実施の形態）

次に、本発明の第３の実施の形態について図面を参照して詳細に説明する。なお、本実施の形態の説明において参照する各図面において、本発明の第１および第２の実施の形態と同一の構成および同様に動作するステップには同一の符号を付して本実施の形態における詳細な説明を省略する。

[0090] まず、本発明の第３の実施の形態としてのメッセージ分析装置３の機能ブロック構成を図８に示す。図８において、メッセージ分析装置３は、本発明の第２の実施の形態としてのメッセージ分析装置２に対して、さらに、クラスタ細分化部３５を備える点が異なる。なお、メッセージ分析装置３およびその各機能ブロックは、図２を参照して説明した本発明の第１の実施の形態と同一のハードウェア要素によって構成可能である。ただし、メッセージ分析装置３および各機能ブロックのハードウェア構成は、上述の構成に限定されない。

[0091] クラスタ細分化部３５は、クラスタリング部２１によって生成されたクラスタ内のメッセージ群を、変数部分の重要性に基づいてさらに分割したクラスタを生成する。このとき、クラスタ細分化部３５は、変数部分の重要性を、変数部分を構成するフィールドの値の品詞に基づいて判定する。詳細には、クラスタ細分化部３５は、変数部分を構成するフィールドの値が文字列として所定の品詞である場合には、そのフィールドが重要であるものとして、

その値の違いに基づいてクラスタを細分化する。

[0092] 具体的には、クラスタ細分化部３５は、クラスタ内において少なくとも１つのメッセージにおいて値が変異しているフィールドを特定する。そして、クラスタ細分化部３５は、特定したフィールドのとる値の文字列としての品詞が、所定の品詞であるか否かに基づいて、そのフィールドの重要性を判定する。なお、クラスタ細分化部３５は、特定したフィールドについて、クラスタ内のいずれかのメッセージ（例えば、代表メッセージ）における値に基づき品詞の判定を行ってもよい。また、クラスタ細分化部３５は、文字列（単語）の品詞を記憶した辞書を用いて、品詞の判定を行えばよい。そのような辞書は、例えば、メモリ１００２にあらかじめ記憶されていてもよい。所定の品詞としては、例えば、動詞、副詞、および、形容詞等が定められる。

[0093] なお、クラスタの細分化により、分割前のクラスタでは変数部分として識別されていたフィールドのうち、重要であると判定されたフィールドは、分割後のクラスタでは不変部分として識別されることになる。

[0094] 以上のように構成されたメッセージ分析装置３の動作について、図９を参照して説明する。

[0095] まず、メッセージ分析装置３は、ステップＳ２１～Ｓ２４まで本発明の第２の実施の形態と同様に動作して、形成したクラスタについてフィールドの解析を行い、全体の類似性が所定条件を満たすか否かを判定する。

[0096] ここで、全体の類似性が所定条件を満たすと判定されたクラスタについて、クラスタ細分化部３５は、変数部分のフィールドの値の品詞に基づいて、さらに細分化を行う（ステップＳ３５）。

[0097] 具体的には、前述のように、クラスタ細分化部３５は、変数部分を構成するフィールドの値が文字列であり所定の品詞である場合にはそのフィールドが重要であるものと判定する。そして、クラスタ細分化部３５は、そのフィールドの値の違いに基づいて、クラスタを細分化する。

[0098] 次に、パターン生成部２３は、細分化されたクラスタ、および、細分化されなかったクラスタのそれぞれについて、本発明の第２の実施の形態と同様

にステップＳ２５～Ｓ２７を実行する。ただし、パターン生成部２３は、細分化されたクラスタについて、細分化の基準となったフィールドについては、不変部分としてその値を共通パターンに含める。これにより、パターン生成部２３は、必要に応じて細分化された各クラスタについて、共通パターンおよびその変数部分の引数のリストを、メッセージパターンとして生成し出力する。

[0099] 以上で、メッセージ分析装置３は、動作を終了する。

[0100] 次に、メッセージ分析装置３の動作を具体例で示す。

[0101] ここでは、クラスタリング部２１により図６に示すクラスタＡおよびクラスタＢが生成され、クラスタＡについて図７に示すフィールド解析結果が生成されているものとする（ステップＳ２１～Ｓ２４）。

[0102] 次に、クラスタ細分化部３５は、クラスタの細分化を行う（ステップＳ３５）。

[0103] 具体的には、まず、クラスタ細分化部３５は、図７において変数部分であるフィールドＩＤ３、７、９、１０のうち、代表メッセージＭｓｇ１１３４におけるフィールドＩＤ９の値「started」が所定の品詞（動詞）であると判定する。すなわち、クラスタ細分化部３５は、フィールドＩＤ９は、変異する重要なフィールドであると判定する。

[0104] 一方、クラスタ細分化部３５は、代表メッセージＭｓｇ１１３４におけるフィールドＩＤ３の値「ｈｏｓｔ０３」およびフィールドＩＤ７の値「３５７１」は、所定の品詞（動詞、副詞および形容詞）のいずれでもないと判定する。すなわち、クラスタ細分化部３５は、フィールドＩＤ３およびフィールドＩＤ７は、変異する補助的なフィールドであると判定する。

[0105] そこで、クラスタ細分化部３５は、重要なフィールドであるフィールドＩＤ９の値に基づいて、クラスタＡを細分化する。図１０に、クラスタＡが細分化されたクラスタＡ１～Ａ３を示す。図１０に示すように、クラスタ細分化部３５は、クラスタＡに含まれているメッセージ群のうち、フィールドＩＤ９の値が「ｓｔａｒｔｅｄ」であるＭｓｇ１１３４、Ｍｓｇ１２１１を、

クラスタＡ１に分類する。また、クラスタ細分化部３５は、フィールドＩＤ９の値が「ｓｔｏｐｐｅｄ」であるＭｓｇ２０９１を、クラスタＡ２に分類する。また、クラスタ細分化部３５は、フィールドＩＤ９の値が「ｔｅｒｍｉｎａｔｅｄ」であるＭｓｇ４６２５を、クラスタＡ３に分類する。

[0106] また、クラスタ細分化部３５は、クラスタＢについても同様に、変数部分のフィールドの値の品詞に基づいて分割を行い、細分化されたｎ個のクラスタＢ１～Ｂｎ（ｎは１以上の整数）を生成したものとする。

[0107] 次に、パターン生成部２３は、細分化されたクラスタＡ１～Ａ３およびクラスタＢ１～Ｂｎについて、メッセージパターンを生成する（ステップＳ２５～Ｓ２７）。

[0108] 例えば、クラスタＡ１については、共通パターン「（Ｄａｔｅ） （Ｔｉｍｅ） ＊ ｐｒｏｃｅｓｓ ａｂｃ ［ ＊ ］ ｓｔａｒｔｅｄ」が生成される。また、さらに、クラスタＡ１については、変数部分であるフィールドＩＤ３の引数のリスト「ｈｏｓｔ０３、ｈｏｓｔ０２」や、フィールドＩＤ７の引数のリスト「３５７１、２２６９」が生成される。

[0109] また、クラスタＡ２については、共通パターン「（Ｄａｔｅ） （Ｔｉｍｅ） ｈｏｓｔ０２ ｐｒｏｃｅｓｓ ａｂｃ ［ ２２６９ ］ ｓｔｏｐｐｅｄ ａｂｎｏｒｍａｌｌｙ」が生成される。

[0110] また、クラスタＡ３については、共通パターン「（Ｄａｔｅ） （Ｔｉｍｅ） ｈｏｓｔ０３ ｐｒｏｃｅｓｓ ａｂｃ ［ ３５７１ ］ ｔｅｒｍｉｎａｔｅｄ」が生成される。

[0111] このように、パターン生成部２３は、クラスタＡ１～Ａ３において、分割の基準となったフィールドＩＤ９については、不変部分としてその値を共通パターンに含めている。また、この例では、クラスタＡ２およびＡ３において、分割前のクラスタＡにおいて変数部分であったフィールドＩＤ３、７、１０についてもそれぞれ値が同一となっている。そのため、パターン生成部２３は、クラスタＡ２およびＡ３の共通パターンに、フィールドＩＤ３、７、１０の値を含めている。ただし、パターン生成部２３は、クラスタ細分化

部３５によって重要でないと判定された変数部分のフィールドについては、分割後のクラスタにおいて値が一致していない場合には、変数部分として共通パターンを生成することになる。

[0112] パターン生成部２３は、クラスタＢ１～Ｂｎについても、同様にしてメッセージパターンを生成する。

[0113] 以上で、具体例の説明を終了する。

[0114] 次に、本発明の第３の実施の形態の効果について述べる。

[0115] 本発明の第３の実施の形態としてのメッセージ分析装置は、大量のメッセージをより少数のメッセージパターンの集合として提示する際に、メッセージ中で重要な情報の内容や傾向を、利用者により的確に把握させることができる。

[0116] その理由について述べる。本実施の形態では、本発明の第２の実施の形態と同様の構成に加えて、クラスタ細分化部が、クラスタに含まれるメッセージ群を、変数部分のフィールドの重要性に基づいてさらに細分化するからである。そして、パターン生成部が、細分化されたクラスタについて、メッセージパターンを生成するからである。

[0117] これにより、本実施の形態は、重要な変数部分の値をメッセージパターンに明示的に含め、補助的な変数部分の値を含めない。つまり、本実施の形態は、変異する部分のうち、主要な情報と補助的な情報とを区別することができる。その結果、本実施の形態は、変異する部分であっても主要な情報についてはその値をそのままメッセージパターンに反映させることができる。

[0118] さらに、本発明の第３の実施の形態としてのメッセージ分析装置は、システムから出力される大量のメッセージをより少数のメッセージパターンの集合として提示する際に、システムの振舞いや状況等の内容や傾向を、利用者により的確に把握させることができる。

[0119] ここで、システムによって記録される大量のメッセージ群の分析をおこなう分析者は、システム上でなにが起きているかを、メッセージ群から推測しなくてはならない。しかしながら、システムの振る舞いや状況を表すフィー

ルドの部分が変数として認識されてしまうとメッセージパターンにその値が現れない。例えば、メッセージ中における動詞や副詞、形容詞といった品詞の部分は、システムの動作や状態を表し重要な意味を持つ可能性が高い。そのような部分の値がメッセージパターンに含まれていないと、解析者にとって、システムの動作や状況を把握しづらくなる。

[0120] 本実施の形態は、メッセージ中の変数部分のフィールドの値が所定の品詞（動詞、副詞および形容詞等）であれば、そのフィールドの値に基づいてクラスタを細分化する。これにより、本実施の形態は、メッセージ中において、システムの動作や状態等を表す重要な情報をそのままメッセージパターンに反映することになる。その結果、本実施の形態を利用する解析者は、メッセージ群の出力元のシステムの振舞いや状況等の重要な情報を、メッセージパターンに基づき正しく把握することが可能となる。

[0121] （第４の実施の形態）

次に、本発明の第４の実施の形態について図面を参照して詳細に説明する。なお、本実施の形態の説明において参照する各図面において、本発明の第１から第３の実施の形態と同一の構成および同様に動作するステップには同一の符号を付して本実施の形態における詳細な説明を省略する。

[0122] まず、本発明の第４の実施の形態としてのメッセージ分析装置４の機能ブロック構成を図１１に示す。図１１において、メッセージ分析装置４は、本発明の第３の実施の形態としてのメッセージ分析装置３に対して、クラスタ細分化部３５に替えてクラスタ細分化部４５を備える点が異なる。

[0123] クラスタ細分化部４５は、本発明の第３の実施の形態におけるクラスタ細分化部３５と略同様に、クラスタリング部２１によって生成されたクラスタ内のメッセージ群を、変数部分の重要性に基づいてさらに分割したクラスタを生成する。ただし、クラスタ細分化部４５は、本発明の第３の実施の形態におけるクラスタ細分化部３５に対して、変数部分の重要性を、変数部分を構成するフィールド間の相関に基づいて判定する点が異なる。

[0124] 詳細には、クラスタ細分化部４５は、変数部分を構成する複数のフィール

ド間に相関が有る場合には、それらのフィールドが重要であるものとして、その値の違いに基づいてクラスタを細分化する。

[0125] 具体的には、クラスタ細分化部４５は、クラスタ内において少なくとも１つのメッセージにおいて値が変異しているフィールドを特定する。そして、クラスタ細分化部４５は、変異しているフィールドの組合せについて、引数間の共起関係を分析する。共起関係があるとは、ある変数（フィールド）の値（引数）と別の変数の値とがあるメッセージに同時に出現することである。

[0126] そして、クラスタ細分化部４５は、そのクラスタ内のメッセージ群において、ある変数の値と別の変数の値とが１対１に対応する場合、各フィールド間に相関があると判定してもよい。また、クラスタ細分化部４５は、変数部分を構成するフィールドの組合せについて、引数間の共起確率を算出してもよい。この場合、クラスタ細分化部４５は、引数間の共起確率がランダムな確率に比べて有意に高い（例えば、閾値以上である）場合に、フィールド間に相関があると判定してもよい。

[0127] また、クラスタ細分化部４５は、相関があると判定した各フィールドを重要であるとして、それらの値に基づきクラスタを細分化する。

[0128] 以上のように構成されたメッセージ分析装置４の動作について、図１２を参照して説明する。

[0129] まず、メッセージ分析装置４は、ステップＳ２１～Ｓ２４まで本発明の第２の実施の形態と同様に動作して、形成したクラスタについてフィールドの解析を行い、全体の類似性が所定条件を満たすか否かを判定する。

[0130] 次に、クラスタ細分化部４５は、全体の類似性が所定条件を満たすと判定されたクラスタについて、変数部分を構成する複数のフィールド間の相関の有無に基づいて、さらに細分化を行う（ステップＳ４５）。

[0131] 具体的には、前述のように、クラスタ細分化部４５は、変数部分を構成する複数のフィールドの組合せについて、それらの引数が１対１に対応する（または引数間の共起確率が閾値以上である）場合には、それらのフィールド

が重要であるものと判定する。そして、クラスタ細分化部４５は、それらのフィールドの値の違いに基づいて、クラスタを細分化する。

[0132] 次に、メッセージ分析装置４は、本発明の第３の実施の形態と同様にステップＳ２５～Ｓ２７までを実行する。これにより、パターン生成部２３は、必要に応じて細分化された各クラスタについて、共通パターンおよびその変数部分の引数のリストを、メッセージパターンとして生成し出力する。

[0133] 以上で、メッセージ分析装置４は、動作を終了する。

[0134] 次に、メッセージ分析装置４の動作を具体例で示す。

[0135] ここでは、クラスタリング部２１により図６に示すクラスタＡおよびクラスタＢが生成され、クラスタＢについて図１３に示すフィールド解析結果が生成されているとする（ステップＳ２１～Ｓ２４）。

[0136] 次に、クラスタ細分化部４５は、フィールド間の相関に基づきクラスタの細分化を行う（ステップＳ４５）。

[0137] 具体的には、まず、クラスタ細分化部４５は、クラスタＢにおいて変数部分であるフィールドＩＤ３、７、１１の組合せについて、引数の共起関係を分析する。共起関係の分析結果を模式的に図１４に示す。図１４において、左図は、フィールドＩＤ３および７間での引数の共起関係を示している。また、右図は、フィールドＩＤ７および１１間での引数の共起関係を示している。また、図１４において、矩形は各フィールドの値を示している。また、矩形間を結ぶ線は、共起関係を表している。

[0138] 図１４に示したように、フィールドＩＤ３および７間では、その値の出現の仕方に規則性がみられない。一方、フィールドＩＤ７および１１間では、その値が１対１に対応している。つまり、フィールドＩＤ７および１１間では、引数間の共起確率が１００％である。

[0139] この場合、クラスタ細分化部４５は、引数間の共起確率が１００％であるフィールドＩＤ７および１１間に相関があるとみなす。これにより、クラスタ細分化部４５は、相関のあるフィールドＩＤ７および１１を重要なフィールドであると判定する。そして、クラスタ細分化部４５は、これらのフィー

ルドＩＤ７および１１の値（引数）に基づいて、クラスタＢを細分化する。図１５に、クラスタＢが細分化されたクラスタＢ１～Ｂ３を示す。図１５に示すように、クラスタ細分化部４５は、クラスタＢに含まれているメッセージ群のうち、フィールドＩＤ７および１１の値の組合せが「１１９７」および「ｒｅｓｅｔ」であるＭｓｇ３２７を、クラスタＢ１に分類する。また、クラスタ細分化部４５は、フィールドＩＤ７および１１の値の組合せが「１１９０」および「ｅｓｔａｂｌｉｓｈｅｄ」であるＭｓｇ３８８およびＭｓｇ８１９を、クラスタＢ２に分類する。また、クラスタ細分化部４５は、フィールドＩＤ７および１１の値の組合せが「１１９９」および「ｂｒｏｋｅｎ」であるＭｓｇ５２１を、クラスタＢ３に分類する。

[0140] また、クラスタ細分化部４５は、クラスタＡについても同様に、変数部分のフィールド間の相関に基づいて分割を行い、細分化されたｍ個のクラスタＡ１～Ａｍ（ｍは１以上の整数）を生成したものとする。

[0141] 次に、パターン生成部２３は、細分化されたクラスタＡ１～ＡｍおよびクラスタＢ１～Ｂ３について、メッセージパターンを生成する（ステップＳ２５～Ｓ２７）。

[0142] 例えば、クラスタＢ１については、共通パターン「（Ｄａｔｅ）　（Ｔｉｍｅ）　ｈｏｓｔ０３　＜　ＮＣ　－　１１９７＞　ｎｅｔｗｏｒｋ　ｃｏｎｎｅｃｔｉｏｎ　ｒｅｓｅｔ」が生成される。

[0143] また、クラスタＢ２については、共通パターン「（Ｄａｔｅ）　（Ｔｉｍｅ）　＊　＜　ＮＣ　－　１１９０＞　ｎｅｔｗｏｒｋ　ｃｏｎｎｅｃｔｉｏｎ　ｅｓｔａｂｌｉｓｈｅｄ」が生成される。また、さらに、クラスタＢ２については、変数部分であるフィールドＩＤ３の引数のリスト「ｈｏｓｔ０１、ｈｏｓｔ０２」が生成される。

[0144] また、クラスタＢ３については、共通パターン「（Ｄａｔｅ）　（Ｔｉｍｅ）　ｈｏｓｔ０２　＜　ＮＣ　－　１１９９＞　ｎｅｔｗｏｒｋ　ｃｏｎｎｅｃｔｉｏｎ　ｂｒｏｋｅｎ」が生成される。

[0145] このように、パターン生成部２３は、クラスタＢ１～Ｂ３において、分割

の基準となったフィールドＩＤ７および１１については、不変部分としてその値を共通パターンに含めている。また、この例では、クラスタＢ１およびＢ３において、分割前のクラスタＢにおいて変数部分であったフィールドＩＤ３についても値が同一となっている。そのため、パターン生成部２３は、クラスタＢ１およびＢ３の共通パターンに、フィールドＩＤ３の値を含めている。ただし、パターン生成部２３は、クラスタ細分化部４５によって重要でないと判定された変数部分のフィールドについては、分割後のクラスタにおいて値が一致していない場合には、変数部分として共通パターンを生成することになる。

[0146] パターン生成部２３は、クラスタＡ１～Ａｍについても、同様にしてメッセージパターンを生成する。

[0147] 以上で、具体例の説明を終了する。

[0148] 次に、本発明の第４の実施の形態の効果について述べる。

[0149] 本発明の第４の実施の形態としてのメッセージ分析装置は、大量のメッセージをより少数のメッセージパターンの集合として提示する際に、メッセージの設計者の意図を示す情報の内容や傾向を、利用者に的確に把握させることができる。

[0150] その理由について述べる。本実施の形態では、本発明の第２の実施の形態と同様の構成に加えて、クラスタ細分化部が、クラスタに含まれるメッセージ群を、変数部分のフィールド間の相関の有無に基づいてさらに細分化するからである。そして、パターン生成部が、細分化されたクラスタについて、メッセージパターンを生成するからである。

[0151] これにより、本実施の形態は、相関のある変数部分の値をメッセージパターンに明示的に含める。つまり、本実施の形態は、変異する部分のうち、相関のある変数部分を主要な情報として、そうでない補助的な情報とを区別することができる。その結果、本実施の形態は、変異する部分であっても、変数間に相関のある主要な情報についてはその値をそのままメッセージパターンに反映させることができる。

[0152] ここで、このような相関のある変数（フィールド）の値は、メッセージの設計者が、なんらかの意図により事前に設計した情報である可能性が高い。たとえば、システムにより出力されるログでは、メッセージの設計者が、エラーメッセージの種類を表すエラーコードや、エラーメッセージの深刻度を表すエラーレベルなどを、共にメッセージに含めるよう設計することが考えられる。このようなメッセージでは、これらのエラーコードおよびエラーレベルをそれぞれ表すフィールド間には、相関が有ることになる。

[0153] このように、本実施の形態は、変数部分であるフィールド間の相関の有無を分析することによって、メッセージの設計者の意図した重要な情報をメッセージパターンに反映することができる。その結果、本実施の形態を利用するメッセージの分析者は、メッセージパターンから、メッセージの設計者の意図を把握することが可能となる。

[0154] なお、上述した本発明の第３および第４の実施の形態において、クラスタ細分化部が、変数部分を構成するフィールドの値の品詞またはフィールド間の相関の有無に基づいて、クラスタを細分化する例について説明した。これに限らず、クラスタ細分化部は、変数部分を構成するフィールドの重要性をその他の情報に基づいて判定し、重要性があると判定したフィールドの値に基づいて細分化を行ってもよい。

[0155] また、上述した本発明の各実施の形態において、メッセージが、ＩＴシステムの要素により出力されたテキスト情報である例を中心に説明したが、メッセージは、他の要素によって出力された情報であってもよい。また、メッセージは、入力装置を介して入力された情報であってもよい。また、メッセージは、テキスト以外の種類の情報を含んでいてもよい。

[0156] また、上述した本発明の各実施の形態において、クラスタリング部が、一致するフィールドの割合を類似度として、または一致しないフィールドの割合を距離として、クラスタリングを行う例について説明した。これに限らず、クラスタリング部は、メッセージ間の類似性を表す情報として算出可能な他の情報に基づいて類似度または距離を算出し、クラスタリングを行っても

よい。

[0157] また、上述した本発明の各実施の形態において、パターン生成部が、不変部分のフィールドの値を表した情報と、変数部分のフィールドを所定の記号で表した情報とを、フィールドの出現順に並べた情報を共通メッセージとして生成する例について説明した。また、パターン生成部が、変数部分のフィールドの取り得る引数のリストを生成する例について説明した。ただし、メッセージパターンの表現形式を限定するものではない。パターン生成部は、クラスタにおいて不変部分を構成するフィールドの値と、変数部分を構成するフィールドのとる引数の値とを認識可能な表現形式であれば、その他の形式でメッセージパターンを生成してもよい。

[0158] また、上述した本発明の各実施の形態において、メッセージ分析装置の各機能ブロックが、記憶装置またはＲＯＭに記憶されたコンピュータ・プログラムを実行するＣＰＵによって実現される例を中心に説明した。これに限らず、各機能ブロックの一部、全部、または、それらの組み合わせが専用のハードウェアにより実現されていてもよい。

[0159] また、上述した本発明の各実施の形態において、メッセージ分析装置の機能ブロックは、複数の装置に分散されて実現されてもよい。

[0160] また、上述した本発明の各実施の形態において、各フローチャートを参照して説明したメッセージ分析装置の動作を、本発明のコンピュータ・プログラムとしてコンピュータの記憶装置（記憶媒体）に格納しておいてもよい。そして、係るコンピュータ・プログラムを当該ＣＰＵが読み出して実行するようにしてもよい。そして、このような場合において、本発明は、係るコンピュータ・プログラムのコードあるいは記憶媒体によって構成される。

[0161] また、上述した各実施の形態は、適宜組み合わせて実施されることが可能である。

産業上の利用可能性

[0162] 本発明は、大量のメッセージの中から、複数のメッセージの共通部分および変異部分を、あらかじめ変異部分を定義する必要なく抽出することができ

、メッセージの内容や傾向の分析を提示する装置として好適である。また、本発明は、システムのログ監視業務において、通知の不要なログをフィルタリングするログ監視ツールにおいて、フィルタリング対象となるメッセージパターンの定義を機械的に生成する装置として好適である。また、本発明は、システム異常時のログ分析作業時において、異常時に大量に発生するエラーメッセージ群から特徴的なログを抽出して分析する作業を支援する装置として好適である。また、本発明は、インターネット上のソーシャルネットワークサービス等にユーザが書き込む大量のメッセージにおいて、ユーザのトレンドや状況把握等の分析を支援する装置として好適である。

[0163] 以上、上述した実施形態を模範的な例として本発明を説明した。しかしながら、本発明は、上述した実施形態には限定されない。即ち、本発明は、本発明のスコープ内において、当業者が理解し得る様々な態様を適用することができる。

[0164] この出願は、２０１５年６月１１日に出願された日本出願特願２０１５－１１８２１７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

[0165] １、２、３、４　　メッセージ分析装置

１１、２１　　クラスタリング部

１２、２２　　フィールド解析部

１３、２３　　パターン生成部

２４　　クラスタ類似性判定部

３５、４５　　クラスタ細分化部

１００１　ＣＰＵ

１００２　メモリ

１００３　出力装置

１００４　入力装置

請求の範囲

[請求項1] １つ以上のフィールドからなるメッセージの集合であるメッセージ群を、前記メッセージ間の類似性に基づいてクラスタに分類するクラスタリング手段と、

前記クラスタ内のメッセージ群を構成する各フィールドについて、その値が変異している変数部分と、変異していない不変部分とを識別するフィールド解析手段と、

前記変数部分および前記不変部分に基づいて、前記クラスタ内のメッセージ群に共通するメッセージパターンを生成するパターン生成手段と、

を備えたメッセージ分析装置。

[請求項2] 前記クラスタを、前記変数部分の重要性に基づいてさらに分割したクラスタを生成するクラスタ細分化手段をさらに備えたことを特徴とする請求項１に記載のメッセージ分析装置。

[請求項3] 前記クラスタ細分化手段は、前記変数部分の重要性を、前記変数部分を構成するフィールドの値の品詞に基づいて判定することを特徴とする請求項２に記載のメッセージ分析装置。

[請求項4] 前記クラスタ細分化手段は、前記変数部分の重要性を、前記変数部分を構成するフィールド間の相関に基づいて判定することを特徴とする請求項２に記載のメッセージ分析装置。

[請求項5] 前記クラスタリング手段は、前記メッセージと、該メッセージに対して類似性が所定条件を満たす他のメッセージとを同一クラスタに分類することを特徴とする請求項１から請求項４のいずれか１項に記載のメッセージ分析装置。

[請求項6] 前記クラスタ内におけるメッセージ群全体の類似性が所定条件を満たすか否かを判定するクラスタ類似性判定手段をさらに備え、

前記パターン生成手段は、前記クラスタ類似性判定手段によって前記全体の類似性が所定条件を満たすと判定されたクラスタについて、

前記メッセージパターンを生成することを特徴とする請求項１から請求項５のいずれか１項に記載のメッセージ分析装置。

[請求項7] 前記クラスタリング手段は、前記各メッセージにおいて所定のフィールドパターンに合致する部分を、前記メッセージ間で互いに類似するフィールドとみなして前記メッセージ群を前記クラスタに分類し、

前記フィールド解析手段は、前記フィールドパターンに合致する値を持つフィールドを、不変部分として識別することを特徴とする請求項１から請求項６のいずれか１項に記載のメッセージ分析装置。

[請求項8] コンピュータ装置を用いて、

１つ以上のフィールドからなるメッセージの集合であるメッセージ群を、前記メッセージ間の類似性に基づいてクラスタに分類し、

前記クラスタ内のメッセージ群を構成する各フィールドについて、その値が変異している変数部分と、変異していない不変部分とを識別し、

前記変数部分および前記不変部分に基づいて、前記クラスタ内のメッセージ群に共通するメッセージパターンを生成するメッセージ分析方法。

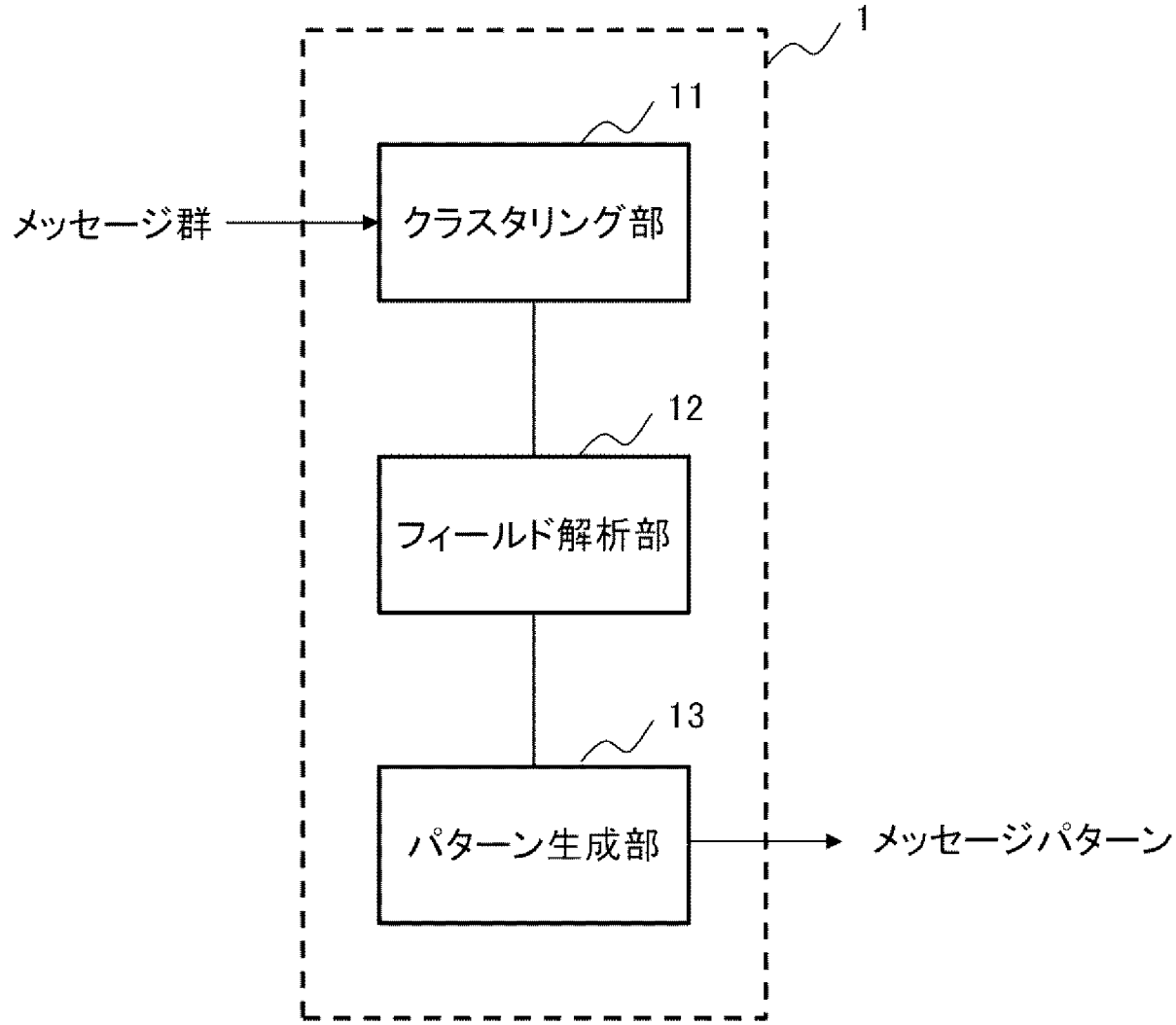
[請求項9] １つ以上のフィールドからなるメッセージの集合であるメッセージ群を、前記メッセージ間の類似性に基づいてクラスタに分類するクラスタリングステップと、

前記クラスタ内のメッセージ群を構成する各フィールドについて、その値が変異している変数部分と、変異していない不変部分とを識別するフィールド解析ステップと、

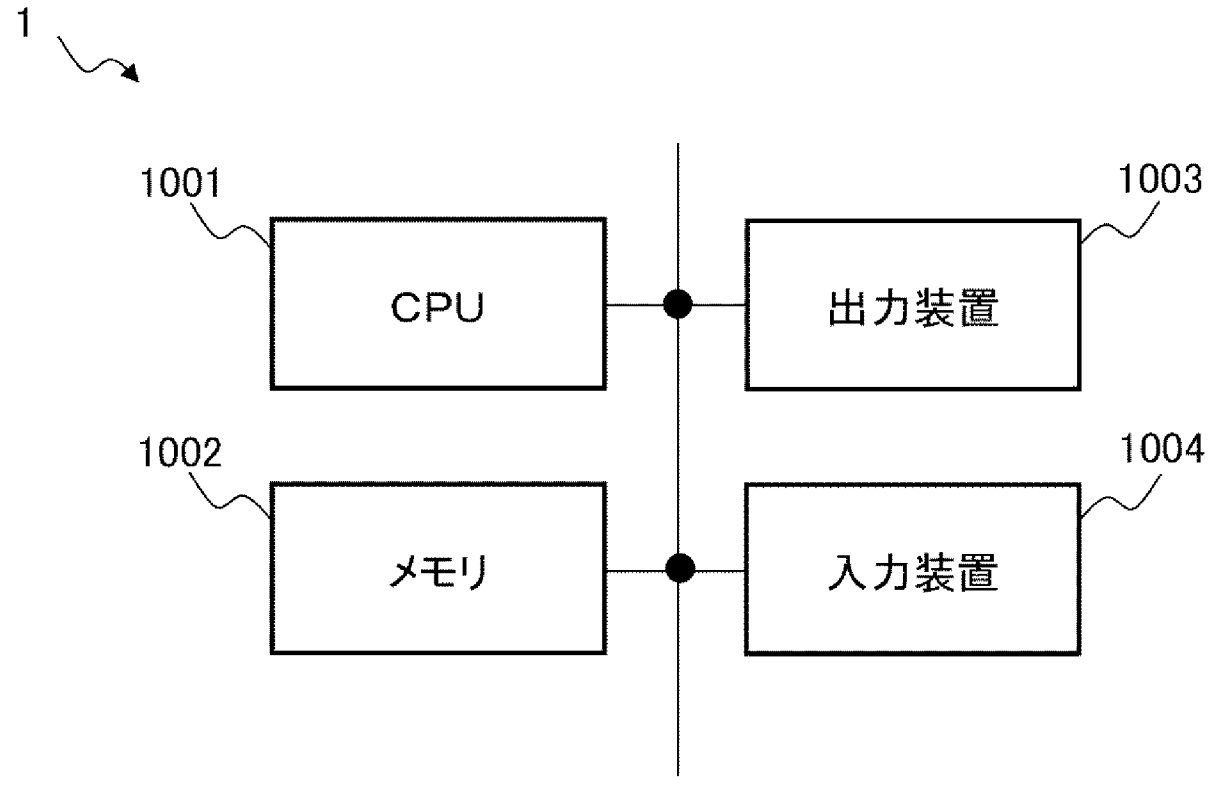
前記変数部分および前記不変部分に基づいて、前記クラスタ内のメッセージ群に共通するメッセージパターンを生成するパターン生成ステップと、

をコンピュータ装置に実行させるメッセージ分析プログラムを記憶した記憶媒体。

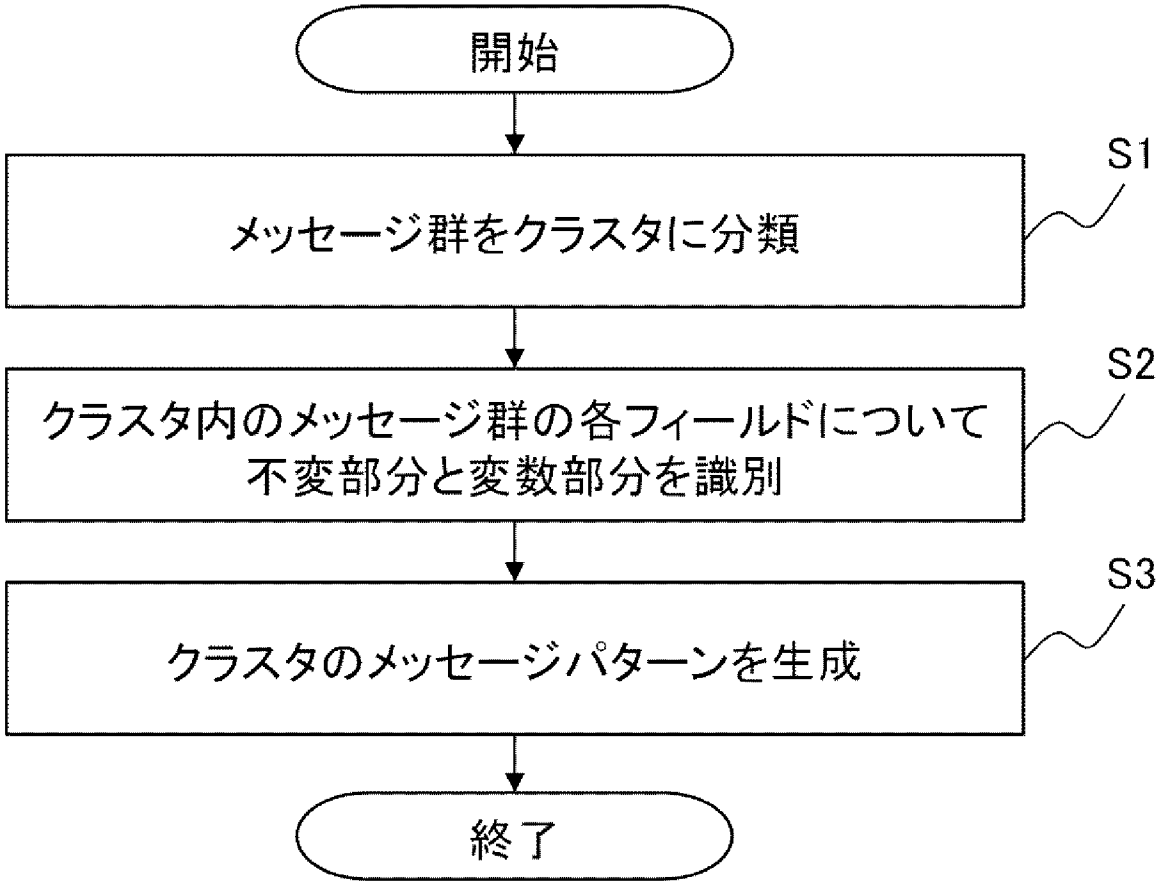
[図1]


1
11
メッセージ群
クラスタリング部
12
フィールド解析部
13
パターン生成部
メッセージパターン

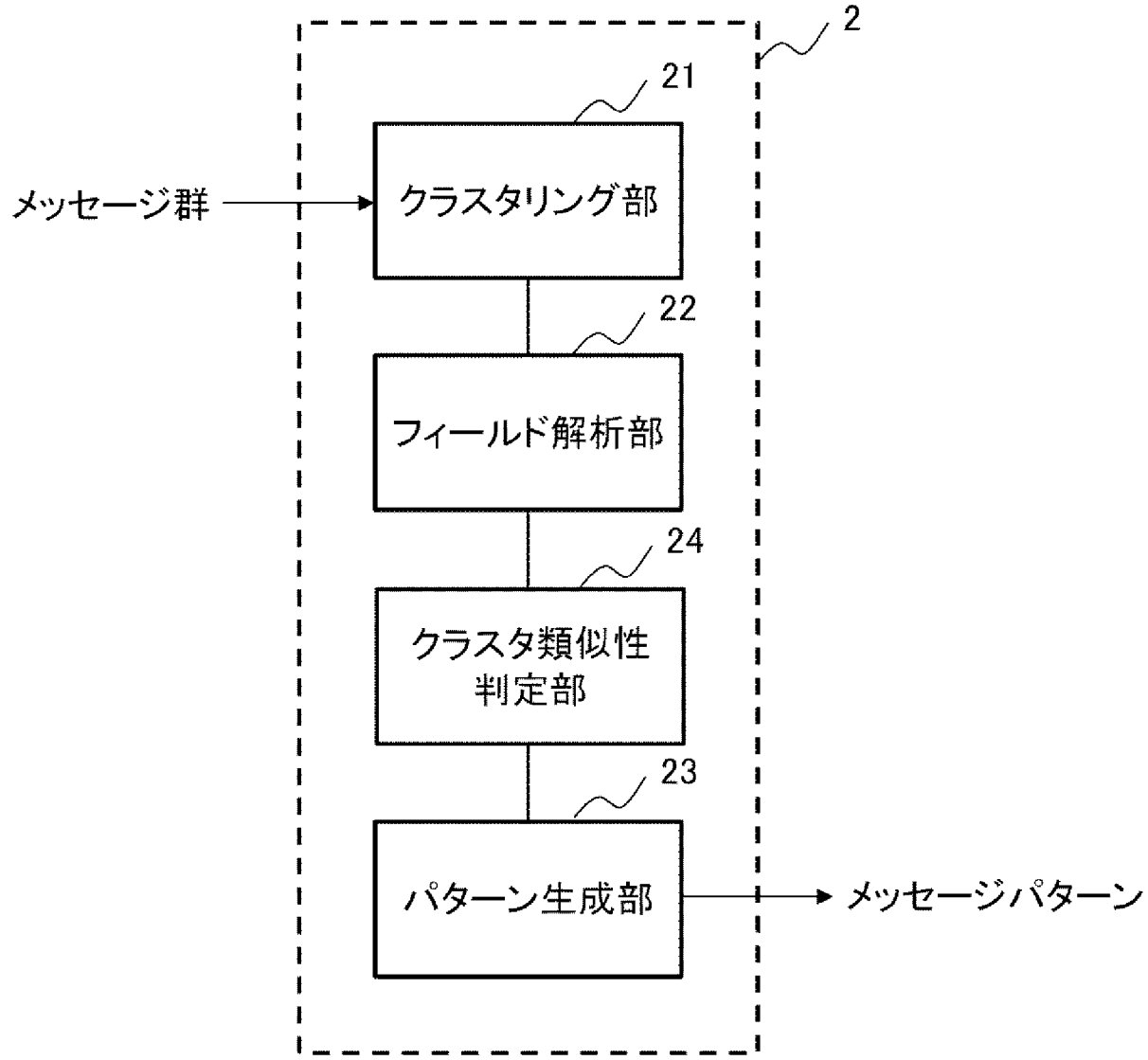
[図2]


1
1001
CPU
1003
出力装置
1002
メモリ
1004
入力装置

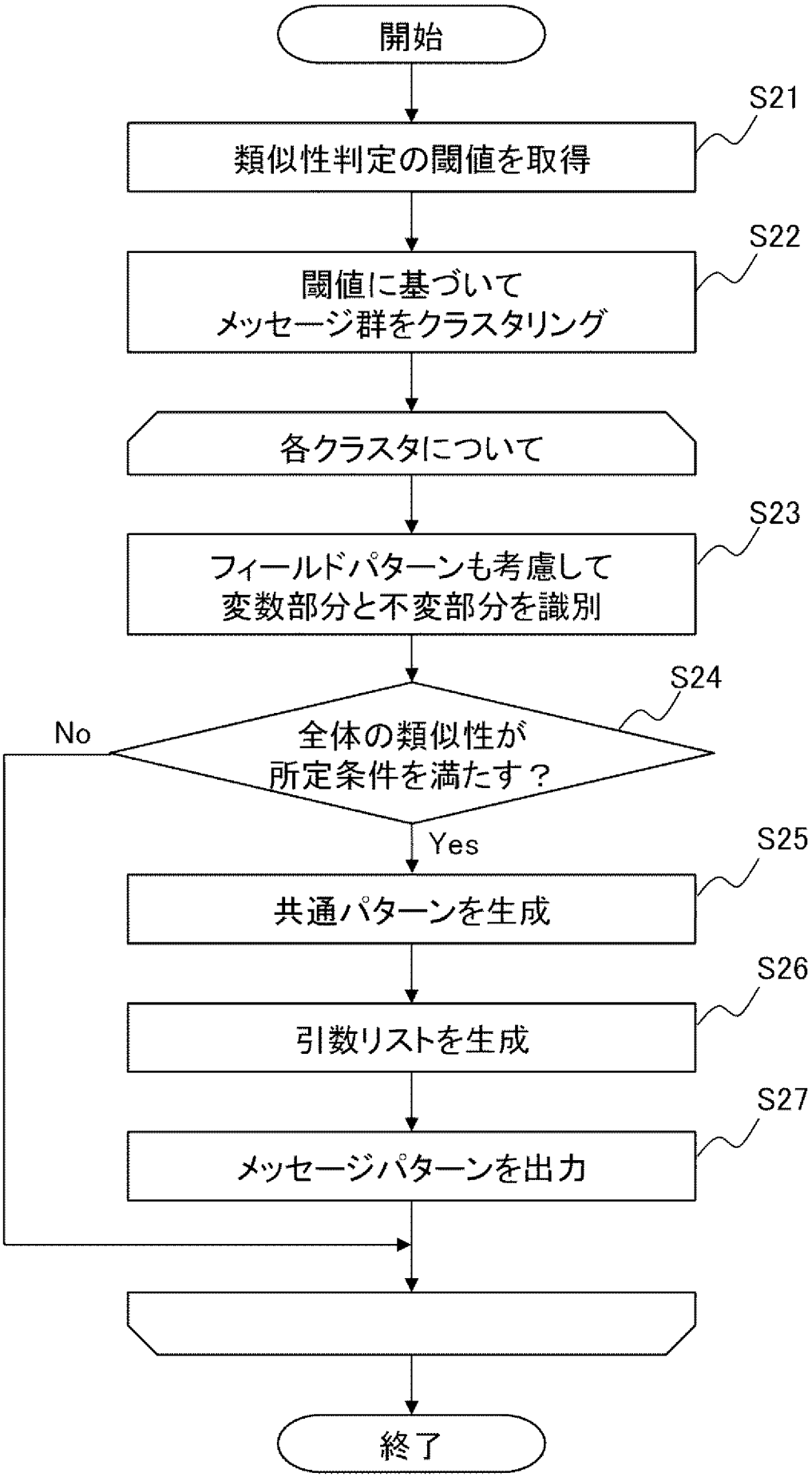
[図3]


開始
S1
メッセージ群をクラスタに分類
S2
クラスタ内のメッセージ群の各フィールドについて
不変部分と変数部分を識別
S3
クラスタのメッセージパターンを生成
終了

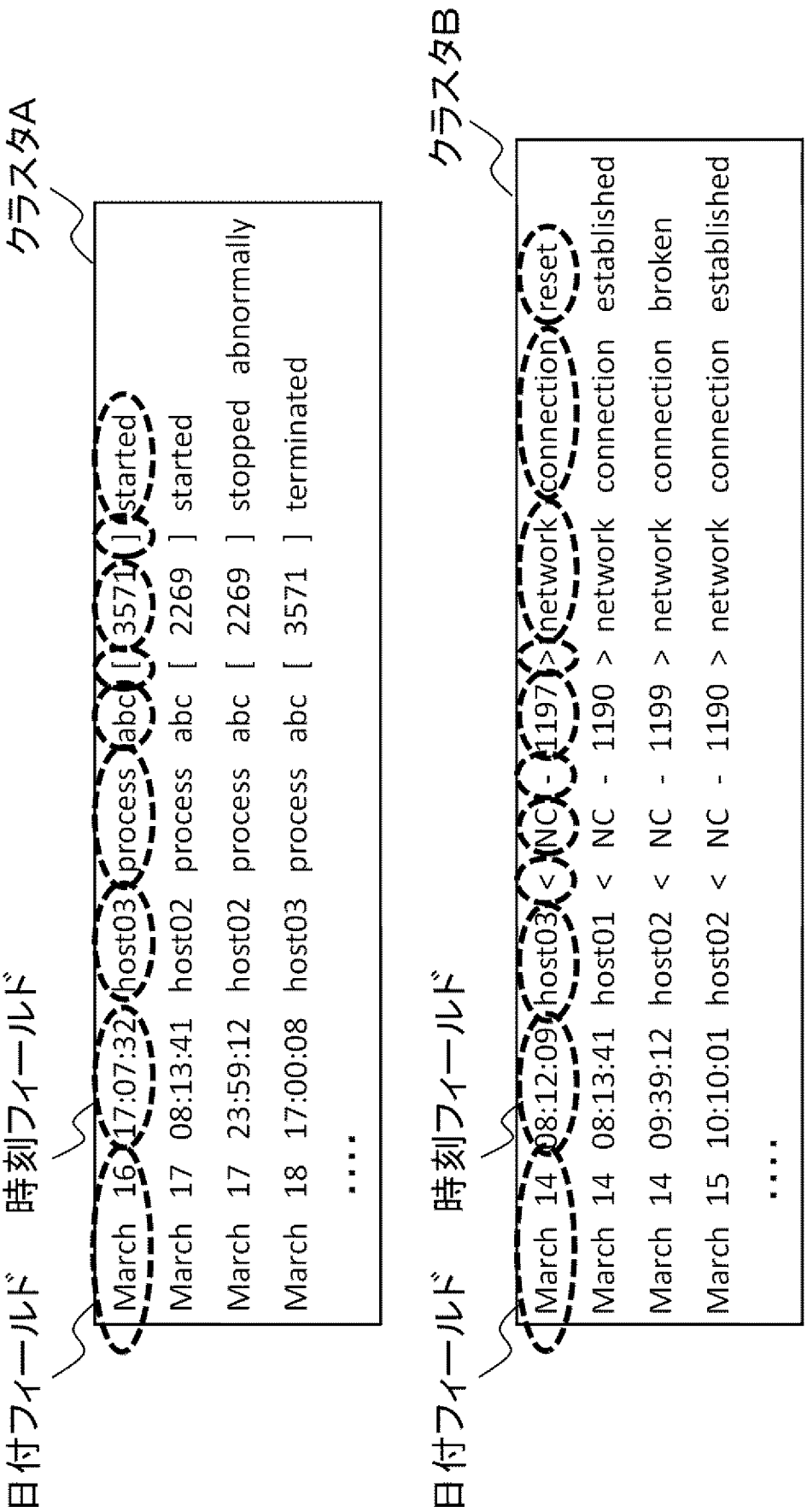
[図4]


2
21
メッセージ群
クラスタリング部
22
フィールド解析部
24
クラスタ類似性
判定部
23
パターン生成部
メッセージパターン

[図5]


開始
S21
類似性判定の閾値を取得
S22
閾値に基づいて
メッセージ群をクラスタリング
各クラスタについて
S23
フィールドパターンも考慮して
変数部分と不変部分を識別
S24
No
全体の類似性が
所定条件を満たす？
Yes
S25
共通パターンを生成
S26
引数リストを生成
S27
メッセージパターンを出力
終了

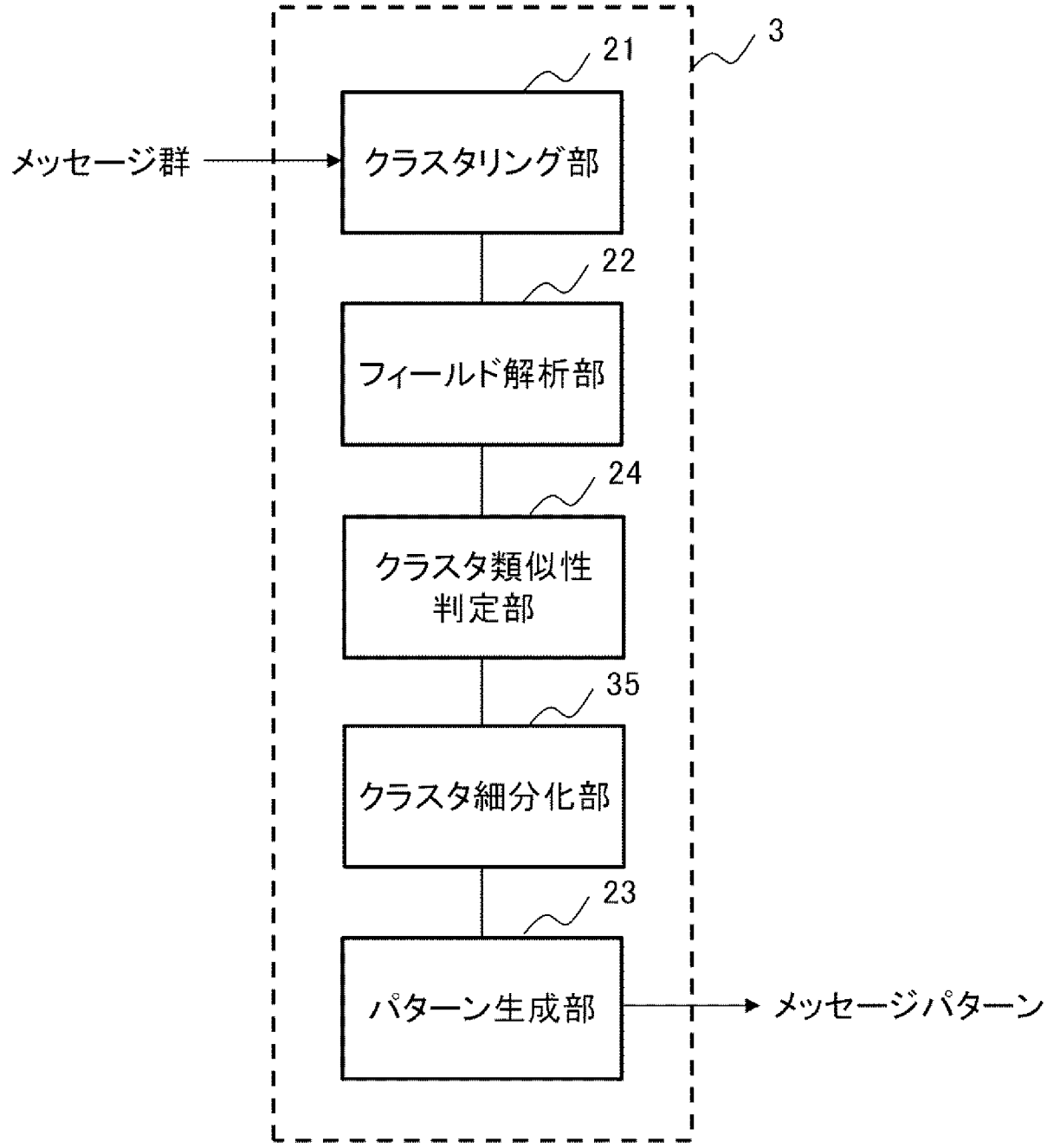
[図6]


クラスタA
日付フィールド
時刻フィールド
March 16 17:07:32 host03 process abc [3571] started
March 17 08:13:41 host02 process abc [2269] started
March 17 23:59:12 host02 process abc [2269] stopped abnormally
March 18 17:00:08 host03 process abc [3571] terminated
・・・・
クラスタB
日付フィールド
時刻フィールド
March 14 08:12:09 host03 < NC - 1197 > network connection reset
March 14 08:13:41 host01 < NC - 1190 > network connection established
March 14 09:39:12 host02 < NC - 1199 > network connection broken
March 15 10:10:01 host02 < NC - 1190 > network connection established
・・・・

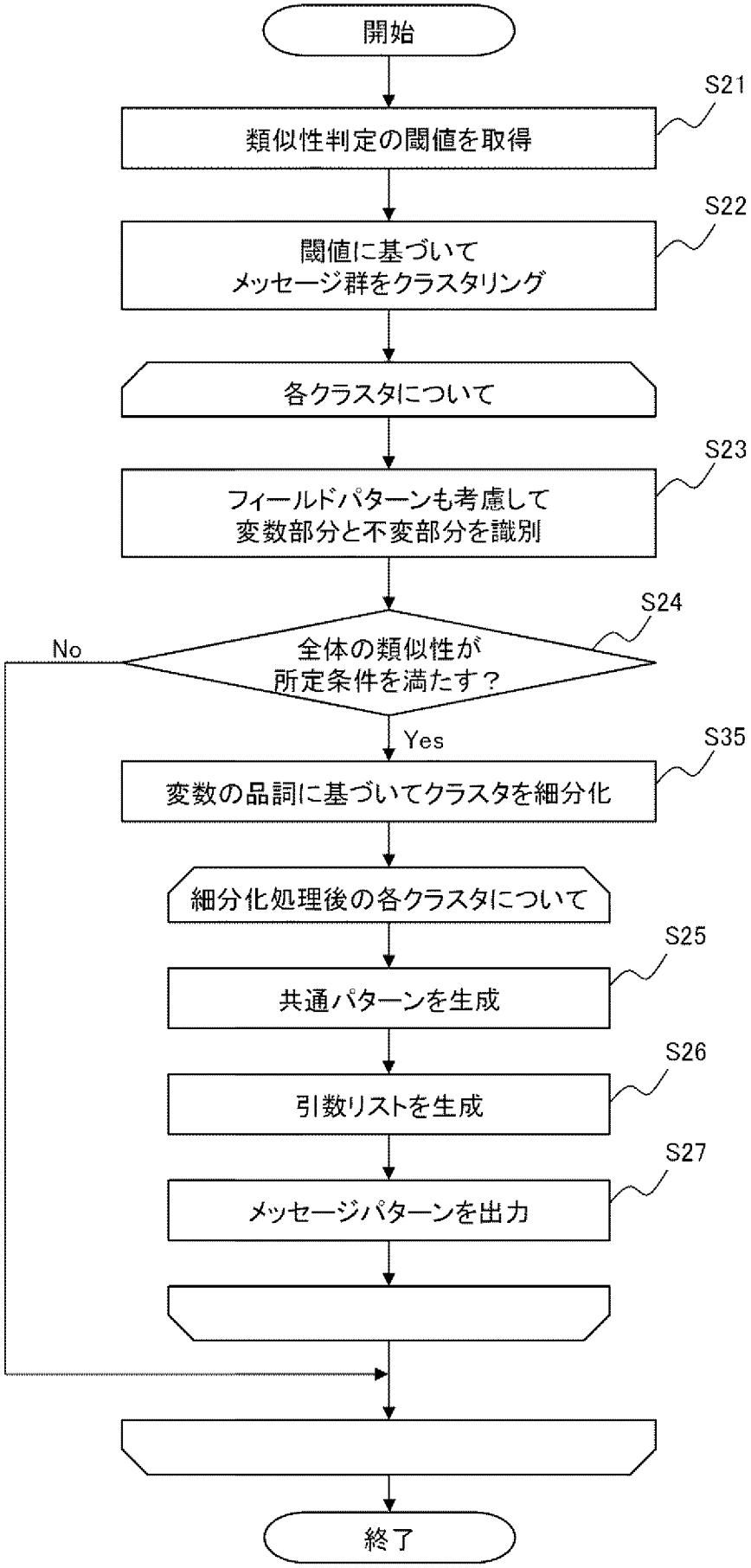
[図7]

フィールドID	Msg1134	Msg1211	Msg2091	Msg4625
1	(Date)			
2	(Time)			
3	host03	host02	host02	
4	process			
5	abc			
6	[			
7	3571	2269	2269	
8	]			
9	started		stopped	terminated
10			abnormally	

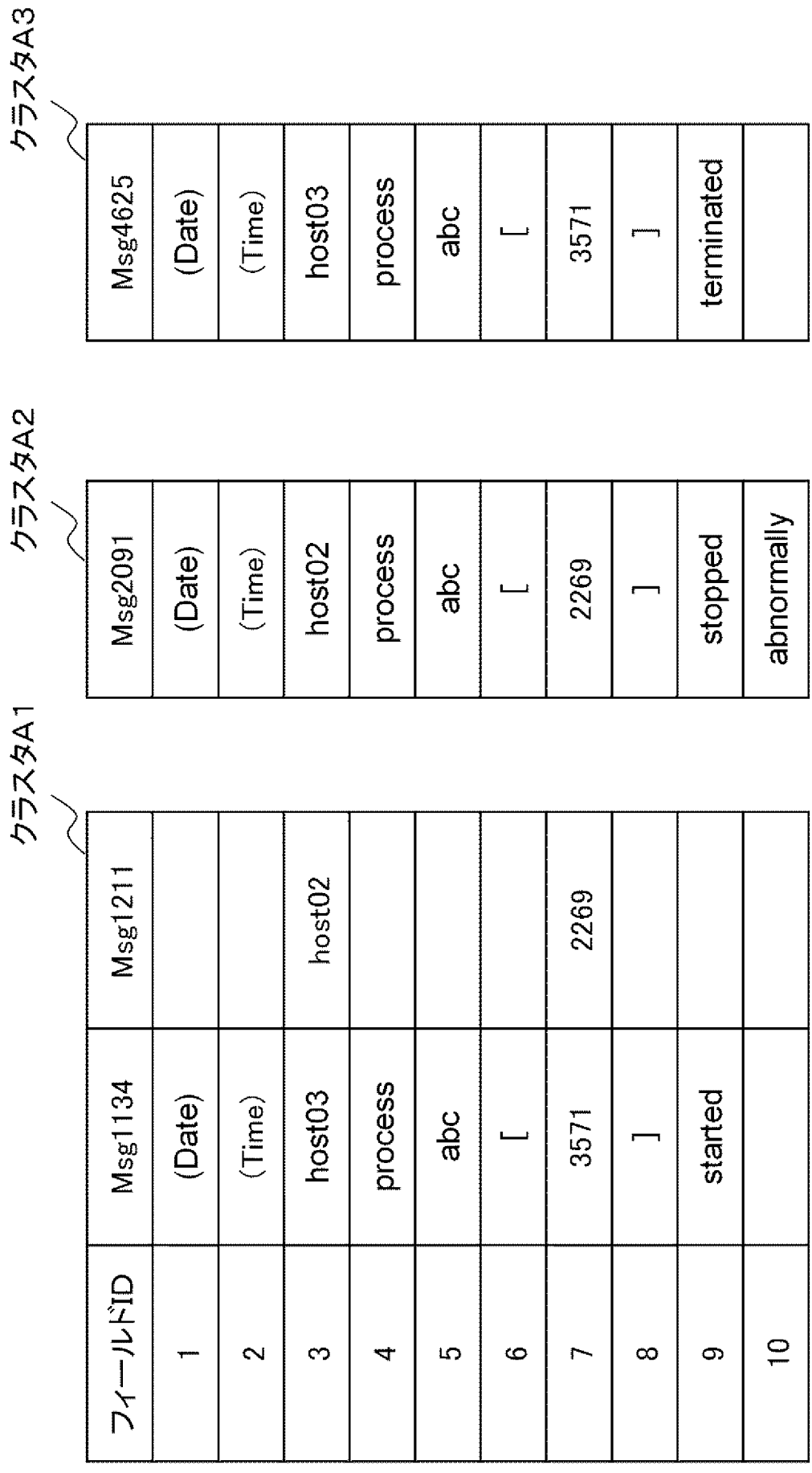
[図8]


3
21
メッセージ群
クラスタリング部
22
フィールド解析部
24
クラスタ類似性
判定部
35
クラスタ細分化部
23
パターン生成部
メッセージパターン

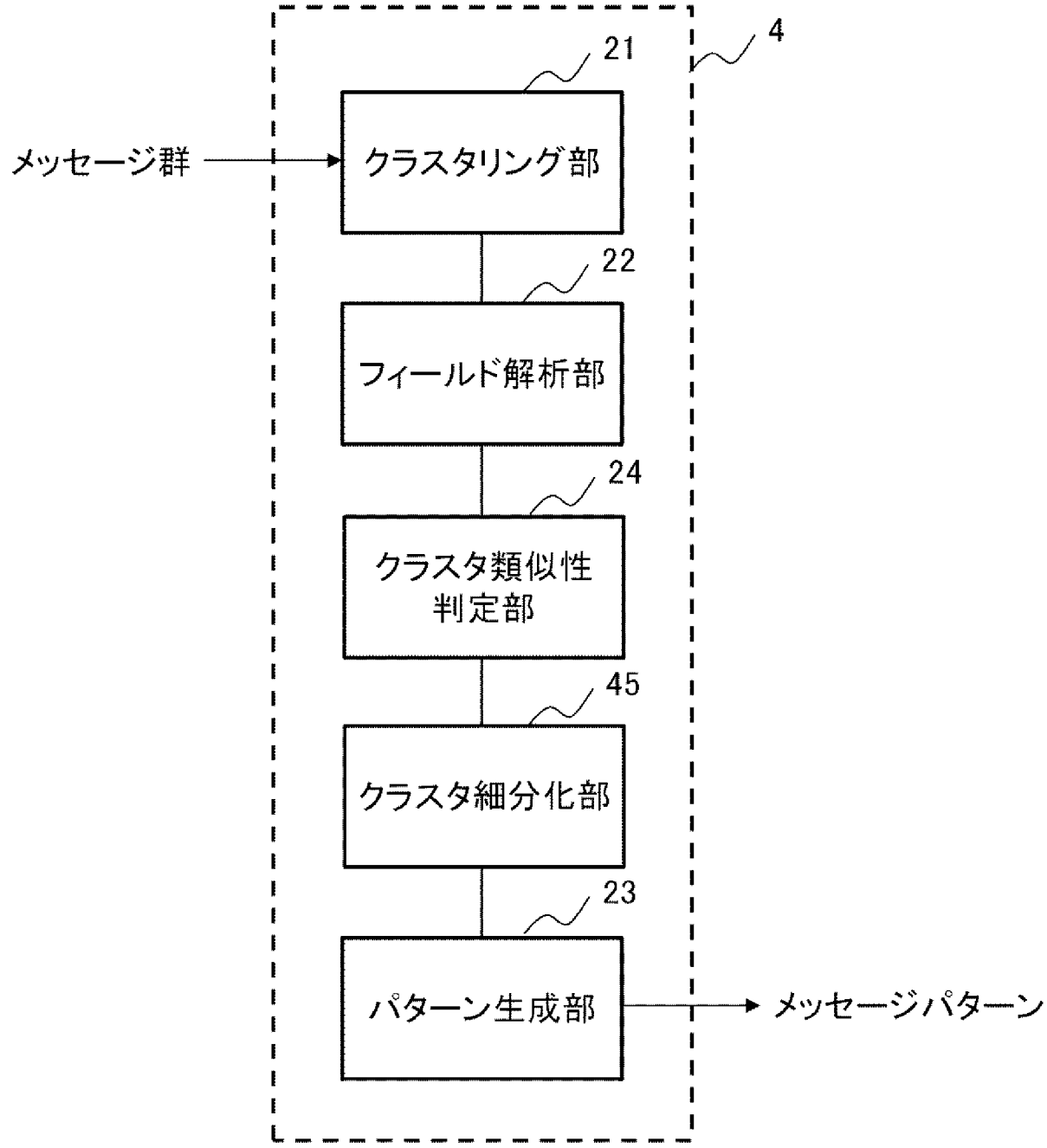
[図9]


開始
S21
類似性判定の閾値を取得
S22
閾値に基づいて
メッセージ群をクラスタリング
各クラスタについて
S23
フィールドパターンも考慮して
変数部分と不変部分を識別
S24
全体の類似性が
所定条件を満たす？
No
Yes
S35
変数の品詞に基づいてクラスタを細分化
細分化処理後の各クラスタについて
S25
共通パターンを生成
S26
引数リストを生成
S27
メッセージパターンを出力
終了

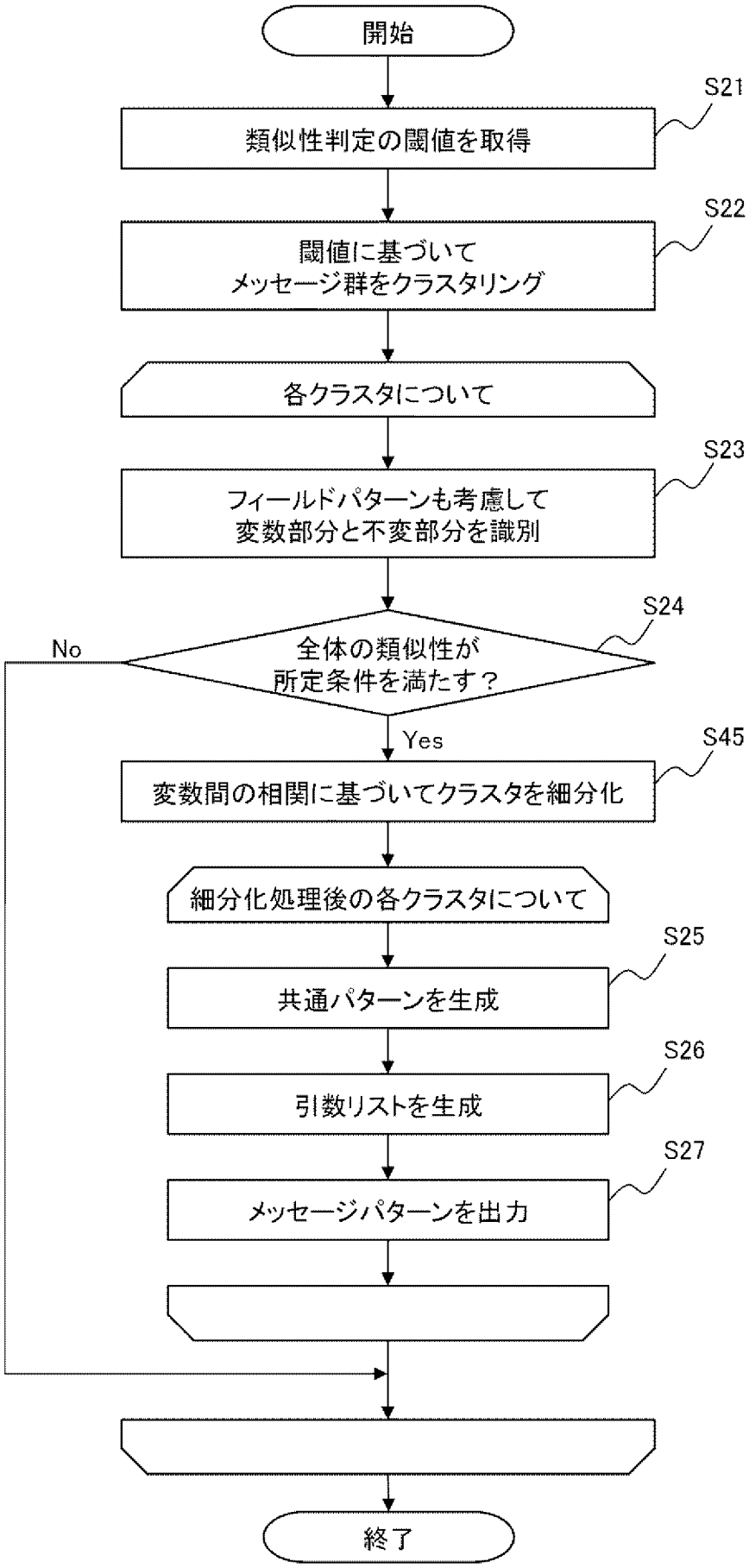
[図10]


クラスタA1
クラスタA2
クラスタA3
フィールドID
Msg1134
Msg1211
Msg2091
Msg4625
(Date)
(Time)
host03
host02
process
abc
[
3571
]
started
stopped
terminated
abnormally
1
2
3
4
5
6
7
8
9
10

[図11]


4
21
メッセージ群
クラスタリング部
22
フィールド解析部
24
クラスタ類似性
判定部
45
クラスタ細分化部
23
パターン生成部
メッセージパターン

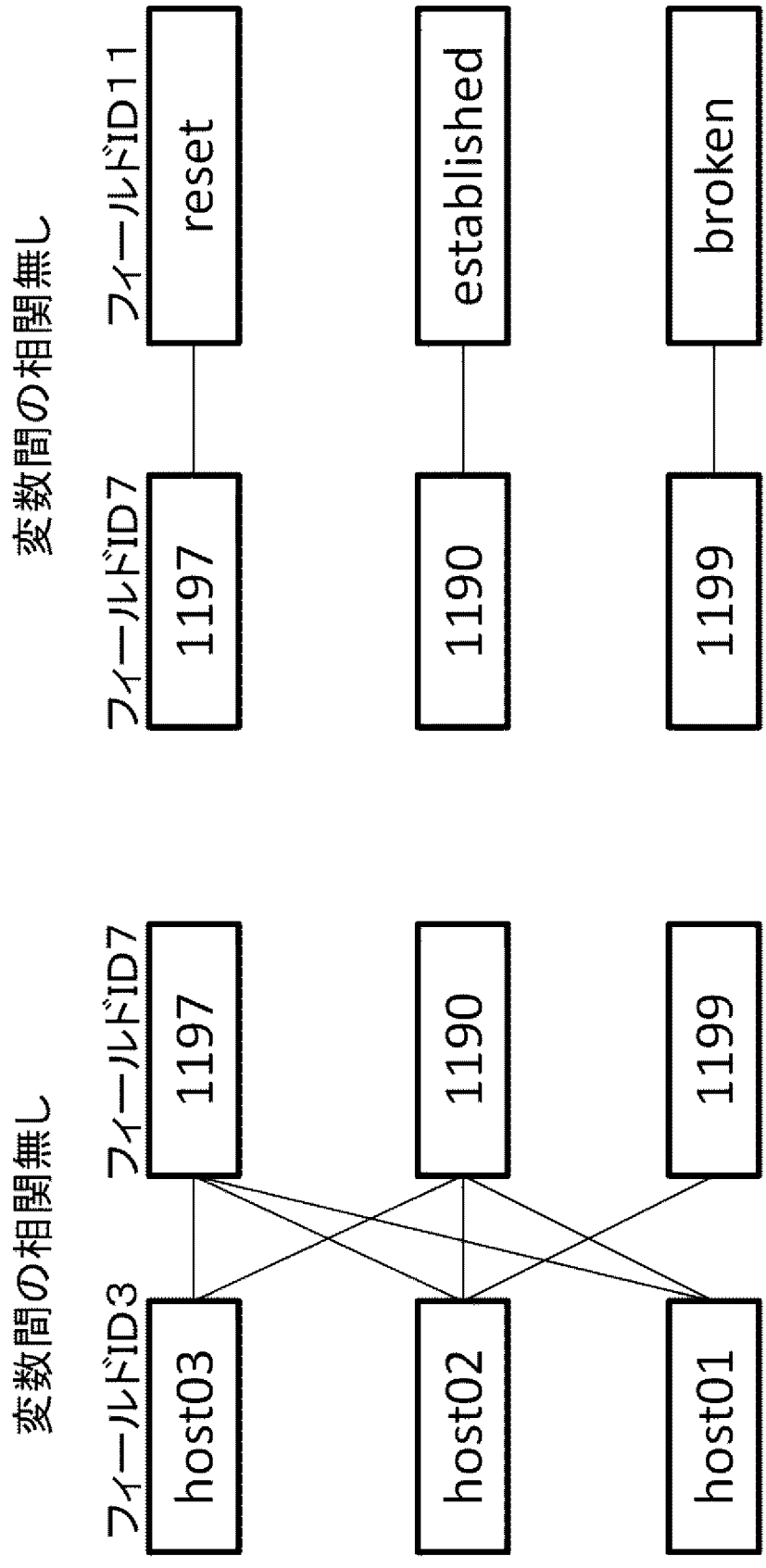
[図12]


開始
S21
類似性判定の閾値を取得
S22
閾値に基づいて
メッセージ群をクラスタリング
各クラスタについて
S23
フィールドパターンも考慮して
変数部分と不変部分を識別
S24
No
全体の類似性が
所定条件を満たす？
Yes
S45
変数間の相関に基づいてクラスタを細分化
細分化処理後の各クラスタについて
S25
共通パターンを生成
S26
引数リストを生成
S27
メッセージパターンを出力
終了

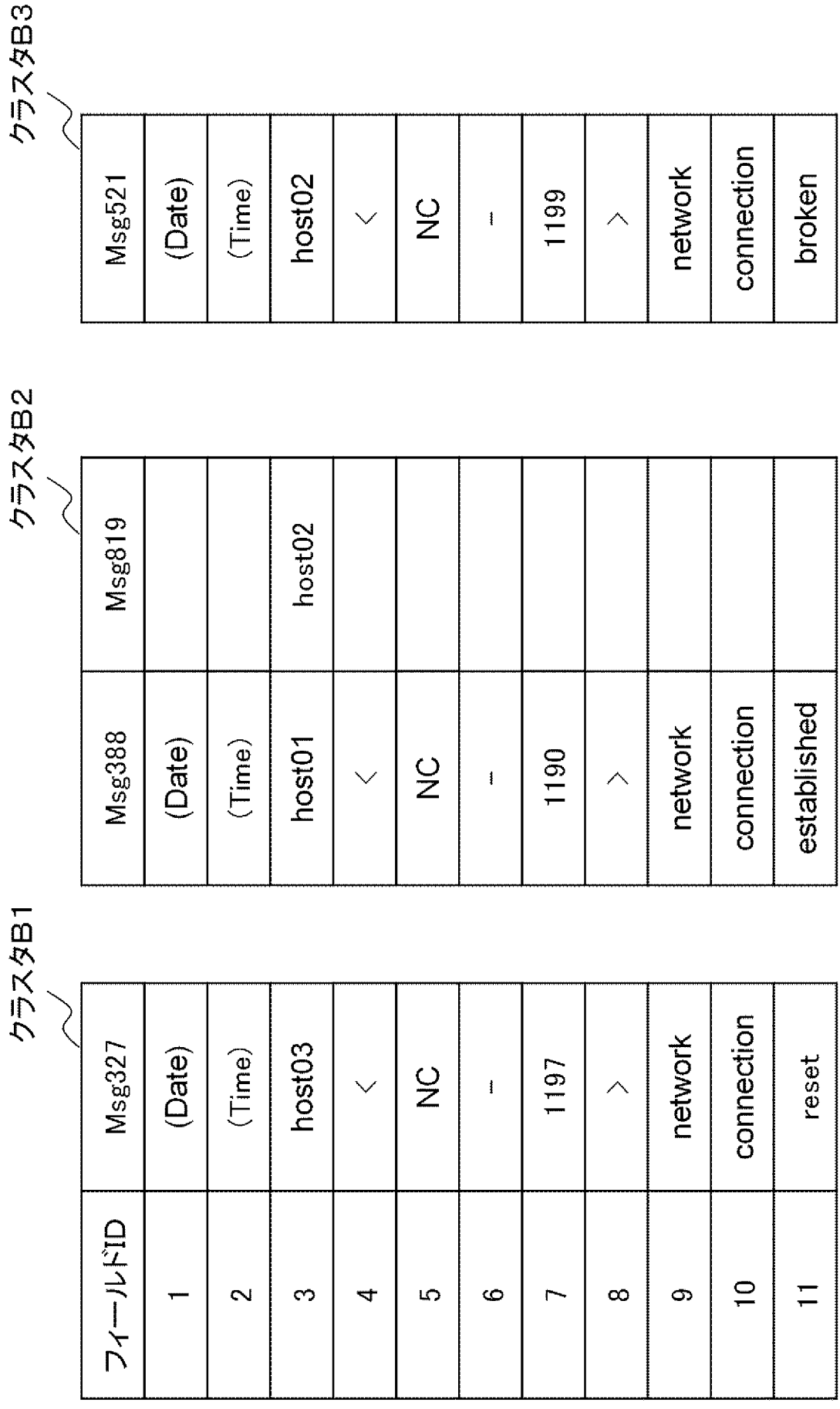
[図13]

フィールドID	Msg327	Msg388	Msg521	Msg819
1	(Date)			
2	(Time)			
3	host03	host01	host02	host02
4	<			
5	NC			
6	-			
7	1197	1190	1199	1190
8	>			
9	network			
10	connection			
11	reset	established	broken	established

[図14]


変数間の相関無し
フィールドID11
フィールドID7
reset
1197
established
1190
broken
1199
変数間の相関無し
フィールドID7
フィールドID3
1197
1190
1199
host03
host02
host01

[図15]


クラスタB1
フィールドID
1
2
3
4
5
6
7
8
9
10
11
Msg327
(Date)
(Time)
host03
<
NC
-
1197
>
network
connection
reset
クラスタB2
Msg388
(Date)
(Time)
host01
<
NC
-
1190
>
network
connection
established
Msg819
host02
クラスタB3
Msg521
(Date)
(Time)
host02
<
NC
-
1199
>
network
connection
broken

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2016/002816

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 1996-2016
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.
Y	JP 2015-036891 A (Nippon Telegraph and Telephone Corp.), 23 February 2015 (23.02.2015), paragraphs [0007] to [0066]; fig. 3 (Family: none)	1-9
Y	JP 2013-214148 A (Toshiba Corp.), 17 October 2013 (17.10.2013), paragraphs [0049] to [0069] (Family: none)	1-9

☒ 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
12 August 2016 (12.08.16)	23 August 2016 (23.08.16)
Name and mailing address of the ISA/ Japan Patent Office 3-4-3,Kasumigaseki,Chiyoda-ku, Tokyo 100-8915,Japan	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2016/002816

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Yuki TSUDA et al., "A Classification Method of Inquiry E-mails for Describing FAQ with Automatic Setting Mechanism of Judgment Thresholds", The Transactions of the Institute of Electrical Engineers of Japan. C, a Publication of Electronics, Information and Systems Society, 01 November 2011 (01.11.2011), vol.131, no.111, pages 1925 to 1932, ISSN 0385-4221, particularly, passage 2-1	2-4,6
Y	Takahiro TANAKA et al., "Consideration on Content Classification of Description Questionnaire Using Self-Organizing Map", IEICE Technical Report NC2010-196 Neurocomputing, 28 February 2011 (28.02.2011), vol.110, no.461, pages 407 to 410, ISSN 0913-5685, particularly, chapter 2	3-4,6
Y	Chinatsu HIROSE et al., "News Clustering Based on Background Similarity", IEICE Technical Report NLC2005-110 Natural Language Understanding and Models of Communication, 26 January 2006 (26.01.2006), vol.105, no.594, pages 25 to 30, ISSN 0913-5685, particularly, passage 3.5	6

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２０１６／００２８１６

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

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（ＩＰＣ））

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	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
Y	JP 2015-036891 A（日本電信電話株式会社）2015.02.23, 段落[0007]-[0066],[図3]（ファミリーなし）	１－９
Y	JP 2013-214148 A（株式会社東芝）2013.10.17, 段落[0049]-[0069]（ファミリーなし）	１－９

☑　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	国際調査報告の発送日
１２．０８．２０１６	２３．０８．２０１６

国際調査機関の名称及びあて先
日本国特許庁（ＩＳＡ／ＪＰ）
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）　５Ｂ　３５６２
多胡　滋
電話番号　０３－３５８１－１１０１　内線　３５４５

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
Ｙ	津田祐輝，外３名，ＦＡＱ作成支援のための分類用閾値自動調節機構を組み入れた問い合わせメール分類方式，電気学会論文誌Ｃ　電子・情報・システム部門誌，2011.11.01，第131巻，第111号，p.1925-1932，ISSN 0385-4221，特に第2-1節	２－４，６
Ｙ	田中崇裕，外２名，自己組織化マップを用いた自由記述形式アンケートの内容分析のための単語クラスタリングに関する検討，電子情報通信学会技術研究報告　ＮＣ２０１０－１９６　ニューロコンピューティング，2011.02.28，第110巻，第461号，p.407-410，ISSN 0913-5685，特に第2章	３－４，６
Ｙ	広瀬千夏，外２名，背景記事集合の類似度に基づく新聞記事のクラスタリング，電子情報通信学会技術研究報告　ＮＬＣ２００５－１１０　言語理解とコミュニケーション，2006.01.26，第105巻，第594号，p.25-30，ISSN 0913-5685，特に第3.5節	６