

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第6201417号  
(P6201417)

(45) 発行日 平成29年9月27日(2017.9.27)

(24) 登録日 平成29年9月8日(2017.9.8)

(51) Int.Cl.

F 1

G06F 17/30 (2006.01)  
G06F 12/00 (2006.01)G06F 17/30 415  
G06F 17/30 414Z  
G06F 12/00 513Z

請求項の数 6 (全 19 頁)

(21) 出願番号	特願2013-105573 (P2013-105573)	(73) 特許権者	000005223 富士通株式会社 神奈川県川崎市中原区上小田中4丁目1番 1号
(22) 出願日	平成25年5月17日(2013.5.17)	(74) 代理人	100089118 弁理士 酒井 宏明
(65) 公開番号	特開2014-228894 (P2014-228894A)	(72) 発明者	森山 泰亘 神奈川県川崎市中原区上小田中4丁目1番 1号 富士通株式会社内
(43) 公開日	平成26年12月8日(2014.12.8)	(72) 発明者	永田 真彦 神奈川県川崎市中原区上小田中4丁目1番 1号 富士通株式会社内
審査請求日	平成28年2月26日(2016.2.26)	(72) 発明者	木野村 光裕 神奈川県川崎市中原区上小田中4丁目1番 1号 富士通株式会社内

最終頁に続く

(54) 【発明の名称】データ抽出プログラム、データ抽出方法およびデータ抽出装置

## (57) 【特許請求の範囲】

## 【請求項 1】

コンピュータに、

複数のキー情報にそれぞれ対応づけられた複数のマスタデータを含むマスタ情報と、前記複数のマスタデータに含まれるデータ情報に関する抽出条件とに基づいて、前記複数のキー情報と前記抽出条件による判定結果とを対応づける判定条件情報を生成し、

ストリームデータの受信に応じ、前記ストリームデータに含まれるキー情報、および、前記判定条件情報に基づき、前記ストリームデータで特定される前記マスタ情報の抽出要否を判定する

処理を実行させるデータ抽出プログラム。

10

## 【請求項 2】

前記生成する処理は、前記キー情報と前記データ情報とを有する前記複数のマスタデータのうち前記データ情報が前記抽出条件を満たすと判定された前記マスタデータのキー情報を対応付けた前記判定条件情報を生成することを特徴とする請求項1に記載のデータ抽出プログラム。

## 【請求項 3】

前記生成する処理は、前記マスタデータのデータ情報が更新された場合に、更新された前記マスタデータのキー情報に対応する判定条件情報の判定結果を、更新後の前記データ情報に基づいた判定結果で更新し、前記マスタデータが削除された場合に、削除された前記マスタデータのキー情報に対応する判定結果を、削除されたことを示す削除情報で更新

20

することを特徴とする請求項 2 に記載のデータ抽出プログラム。

【請求項 4】

前記データ情報が前記抽出条件を満たすか否かを判定するのに際して、前記データ情報を構成する各構成データについて、前記抽出条件を構成する各条件のうち当該構成データに対応する条件を満たすか否かを判定した中間データを生成し、

前記抽出条件が更新された場合に、更新された前記条件に対応する各構成データが当該条件を満たすか否かを判定して前記中間データを更新し、

前記更新した前記中間データと前記抽出条件とに基づいて、前記判定結果を更新する処理をさらにコンピュータに実行させることを特徴とする請求項 2 に記載のデータ抽出プログラム。

10

【請求項 5】

コンピュータが、

複数のキー情報にそれぞれ対応づけられた複数のマスタデータを含むマスタ情報と、前記複数のマスタデータに含まれるデータ情報に関する抽出条件とに基づいて、前記複数のキー情報と前記抽出条件による判定結果とを対応づける判定条件情報を生成し、

ストリームデータの受信に応じ、前記ストリームデータに含まれるキー情報、および、前記判定条件情報に基づき、前記ストリームデータで特定される前記マスタ情報の抽出要否を判定する

処理を含むことを特徴とするデータ抽出方法。

【請求項 6】

20

複数のキー情報にそれぞれ対応づけられた複数のマスタデータを含むマスタ情報と、前記複数のマスタデータに含まれるデータ情報に関する抽出条件とに基づいて、前記複数のキー情報と前記抽出条件による判定結果とを対応づける判定条件情報を生成する生成部と、

ストリームデータの受信に応じ、前記ストリームデータに含まれるキー情報、および、前記判定条件情報に基づき、前記ストリームデータで特定される前記マスタ情報の抽出要否を判定する判定部と

を有することを特徴とするデータ抽出装置。

【発明の詳細な説明】

30

【技術分野】

【0001】

本発明は、データ抽出プログラム、データ抽出方法およびデータ抽出装置に関する。

【背景技術】

【0002】

従来から、受信データが抽出条件に該当するか否かを判定して、データを抽出する技術が利用されている。例えば、データがコード化されたストリームデータを用いて、センサデータやPOS (Point Of Sale system) データなどのデータを収集することが行われている。ストリームデータは、データがコード化されている。このため、ストリームデータとマスタ情報と照合して、データの具体的な内容を抽出することが行われる。

【0003】

40

例えば、マスタ情報は、キーと複数のカラムとから構成されるマスタデータを複数有する。このマスタ情報とストリームデータを用いて、マスタデータを抽出する手法としては、マスタ照合後に抽出条件の判定を行う手法や、マスタデータに対して事前に判定処理を行った結果とストリームデータとを照合する手法とが知られている。

【0004】

例えば、マスタ照合後に抽出条件の判定を行う手法は、ストリームデータ内のコードをキーにして、マスタ情報から該当するマスタデータを特定する。そして、特定したマスタデータの各カラムに対応するデータが、抽出条件を満たすか否かを判定し、抽出条件を満たす場合に、当該マスタデータを抽出する。

【0005】

50

また、事前に判定処理を行う手法は、マスタ情報の各カラムについて、抽出条件を満たすか否かを判定し、カラムごとに条件を満たすキーを抽出した抽出情報を生成する。そして、ストリームデータのコードをキーにして、各カラムの抽出情報を参照して、各カラムの抽出情報にキーが含まれる場合に、当該マスタデータを抽出する。

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開平04-102172号公報

【特許文献2】特開平06-187379号公報

【特許文献3】特開2006-171800号公報

10

【発明の概要】

【発明が解決しようとする課題】

【0007】

しかしながら、上記技術では、例えばカラム数、レコード数、ストリームデータの受信頻度が多くなると、データ抽出にかかる処理時間が増大するという問題がある。

【0008】

例えば、マスタ照合後に抽出条件の判定を行う手法では、マスタデータがM個ある場合には、1つのストリームデータについてM回の判定処理を実行するので、処理時間が長くなる。また、事前に判定処理を行う手法は、カラム数がN個ある場合には、N個の抽出情報を生成することになるので、1つのストリームデータについてN回の判定処理を実行することになり、処理時間が長くなる。

20

【0009】

1つの侧面では、データ抽出を高速化することができるデータ抽出プログラム、データ抽出方法およびデータ抽出装置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】

【0010】

第1の案では、コンピュータに、キー情報とデータ情報を対応付けた各第1のデータにおける前記データ情報が、抽出条件を満たすか否かを判定する処理を実行させる。コンピュータに、判定した前記データ情報に対応付けられる前記キー情報と判定結果とを対応付けた第2のデータを生成する処理を実行させる。コンピュータに、受信データに含まれるキー情報を用いて、各第2のデータから前記判定結果を特定する処理を実行させる。

30

【発明の効果】

【0011】

本発明の1実施形態によれば、データ抽出を高速化することができる。

【図面の簡単な説明】

【0012】

【図1】図1は、実施例1に係るシステムの全体構成例を示す図である。

【図2】図2は、実施例1に係るデータ抽出装置の機能構成を示す機能ブロック図である。

【図3】図3は、マスタ情報DBに記憶される情報の例を示す図である。

40

【図4】図4は、抽出条件式DBに記憶される情報の例を示す図である。

【図5】図5は、中間情報DBに記憶される情報の例を示す図である。

【図6】図6は、真偽判定表の例を示す図である。

【図7】図7は、ストリームデータの例を示す図である。

【図8】図8は、真偽判定表の生成例1を説明する図である。

【図9】図9は、真偽判定表の生成例2を説明する図である。

【図10】図10は、実施例1に係るデータ抽出処理の流れを示すフローチャートである。

。

【図11】図11は、実施例2に係るマスタ情報の削除例を説明する図である。

【図12】図12は、実施例2に係る抽出条件の更新例を説明する図である。

50

【図13】図13は、実施例3に係る真偽判定表更新時のストリームデータ処理例1を説明する図である。

【図14】図14は、実施例3に係る真偽判定表更新時のストリームデータ処理例1を説明する図である。

【図15】図15は、実施例3に係る真偽判定表更新時のストリームデータ処理例2を説明する図である。

【図16】図16は、実施例3に係る真偽判定表更新時のストリームデータ処理例2を説明する図である。

【図17】図17は、真偽判定表の別例を説明する図である。

【図18】図18は、ハードウェア構成例を示す図である。

【発明を実施するための形態】

【0013】

以下に、本願の開示するデータ抽出プログラム、データ抽出方法およびデータ抽出装置の実施例を図面に基づいて詳細に説明する。なお、この実施例によりこの発明が限定されるものではない。また、各実施例は矛盾しない範囲で適宜組み合わせることができる。

【実施例1】

【0014】

【全体構成】

図1は、実施例1に係るシステムの全体構成例を示す図である。図1に示すように、このシステムは、ストリームデータ送信装置1とデータ抽出装置10とが、インターネットなどのネットワーク2を介して接続される。なお、ここでは、ストリームデータを例にして説明するが、これに限定されるものではなく、一般的なデータであっても同様に処理することができる。

【0015】

ストリームデータ送信装置1は、センサデータやPOSデータなどをストリームデータとして、データ抽出装置10に送信する装置の一例である。なお、図1では、1台のストリームデータ送信装置1を図示したが、これに限定されるものではなく、このシステムは複数のストリームデータ送信装置1を有していてもよい。

【0016】

データ抽出装置10は、ストリームデータとマスタ情報と照合して、データの具体的な内容を抽出するサーバ装置の一例である。データ抽出装置10は、キー情報とデータ情報とを対応付けた第1のデータにおけるデータ情報が、抽出条件を満たすか否かを判定する。そして、データ抽出装置10は、判定したデータ情報に対応付けられるキー情報と判定結果とを対応付けた第2のデータを生成する。その後、データ抽出装置10は、受信ストリーミングデータに含まれるキー情報を用いて、各第2のデータから判定結果を特定する。

【0017】

このように、データ抽出装置10は、ストリームデータ受信前に、マスタ情報のデータ抽出判定を行って、キー毎に真偽判定結果を作成しておき、ストリームデータ受信時に、当該データ内のキーと真偽判定結果とを照合することで、データ抽出を高速化する。

【0018】

【データ抽出装置の構成】

図2は、実施例1に係るデータ抽出装置の機能構成を示す機能ブロック図である。図2に示すように、データ抽出装置10は、通信部11、第1記憶部12、第2記憶部13、制御部15を有する。

【0019】

通信部11は、他の装置との通信を確立する処理部であり、例えばネットワークインターフェースカードや無線インターフェースなどである。例えば、通信部11は、ストリームデータ送信装置1が送信したストリームデータを受信する。

【0020】

第1記憶部12は、比較的容量が多いハードディスクなどの記憶装置の一例であり、マスタ情報DB12aと抽出条件式DB12bと中間情報DB12cとを記憶する。また、第1記憶部12は、データ抽出装置10の各種機能を実行するためのデータや各種プログラムを記憶する。

【0021】

マスタ情報DB12aは、ストリームデータで特定される具体的な内容を示すマスタデータを記憶するデータベースである。図3は、マスタ情報DBに記憶される情報の例を示す図である。図3に示すように、マスタ情報DB12aは、「キー、製品、スタイル、価格」を対応付けたマスタデータを記憶する。なお、ここでは、パーソナルコンピュータの製品情報を一例として説明するが、マスタ情報の内容がこれに限定されるものではない。

10

【0022】

ここで記憶される「キー」は、マスタデータを検索する際に使用されるキー情報である。「製品」は、マスタデータが記憶する製品がパーソナルコンピュータな他の製品なのかを示す情報である。「スタイル」は、マスタデータが記憶するパーソナルコンピュータがデスクトップ、タブレット、ノートのいずれであるかを示す情報である。「価格」は、製品の価格を示す情報である。

【0023】

図3の場合、「キー」が「0001」である先頭のマスタデータは、価格が「60000円」である「デスクトップ」型の「PC」であることを示している。「キー」が「0003」である先頭から3番目のマスタデータは、価格が「150000円」である「ノート」型の「PC」であることを示している。

20

【0024】

抽出条件式DB12bは、ストリームデータからマスタデータを抽出する条件式を記憶するデータベースである。ここで記憶される抽出条件式は、管理者等によって設定されて格納される。図4は、抽出条件式DBに記憶される情報の例を示す図である。

【0025】

図4に示す抽出条件は、ストリームデータの「キー」と一致するマスタデータのうち、「製品」が「PC」であり、「スタイル」が「ノート」または「価格」が「100000未満」であるマスタデータを抽出することを示す。

30

【0026】

中間情報DB12cは、マスタデータから真偽判定表を生成する際に生成される中間データを記憶するデータベースである。図5は、中間情報DBに記憶される情報の例を示す図である。図5に示すように、中間情報DB12cは、「キー、製品、スタイル、価格」それぞれについて、抽出条件の判定結果を記憶する。

【0027】

ここで記憶される「キー、製品、スタイル、価格」が示す情報は、図3と同様なので、説明を省略する。図5の場合、キーが「0001」のマスタデータについて、「製品」については「True」と判定され、「スタイル」については「False」と判定され、「価格」については「True」と判定されたことを示す。

【0028】

40

第2記憶部13は、プロセッサが直接読み書きするメモリなどの記憶装置の一例であり、真偽判定表13aを記憶する。真偽判定表13aは、制御部15が第1記憶部12のマスタ情報DB12aを第2記憶部13に展開する際に格納される。

【0029】

図6は、真偽判定表の例を示す図である。図6に示すように、真偽判定表13aは、「キー」と「判定結果」とを対応付けて記憶する。「キー」は、マスタ情報DB12aのマスタデータの「キー」に対応する。「判定結果」は、抽出条件を満たす場合には「1」が格納され、抽出条件を満たさない場合には「0」が格納される。なお、図6では、FBTヒドラの形で生成された真偽判定表を例示したが、これに限定されるものではなく、任意に形式で格納することができる。

50

## 【0030】

制御部15は、データ抽出装置10全体の処理を司る処理部であり、条件判定部16、中間情報生成部17、真偽判定表生成部18、ストリームデータ処理部19を有する。この制御部15は、プロセッサなどの電子回路などの一例である。また、条件判定部16、中間情報生成部17、真偽判定表生成部18、ストリームデータ処理部19は、プロセッサが実行するプロセスの一例である。

## 【0031】

条件判定部16は、マスタデータを構成する各構成データが抽出条件を満たすか否かを判定する処理部である。具体的には、条件判定部16は、マスタ情報DB12aからマスタデータを読み込み、抽出条件式DB12bから抽出条件式を読み込み、各マスタデータを構成する構成データの一例である各カラムが条件を満たすか否かを判定し、その結果を中間情報生成部17に出力する。

10

## 【0032】

例えば、条件判定部16は、マスタ情報DB12aからマスタデータとして「キー(0001)、製品(PC)、スタイル/Desktop)、価格(60000)」を抽出する。また、条件判定部16は、図4示す抽出条件式「製品=PC AND (スタイル=ノート) or (価格=100000未満)」を抽出条件式DB12bから読み込む。

## 【0033】

そして、条件判定部16は、製品が「PC」であることから「製品」のカラムについては条件を満たすと判定する。また、条件判定部16は、スタイルが「Desktop」であることから「スタイル」のカラムについては条件を満たさないと判定する。また、条件判定部16は、価格が「60000」であることから「価格」のカラムについては条件を満たすと判定する。

20

## 【0034】

その後、条件判定部16は、「キー」が「0001」のマスタデータについては「製品(True)、スタイル(False)、価格(True)」であることを中間情報生成部17に出力する。

## 【0035】

中間情報生成部17は、条件判定部16による判定結果に基づいて、中間情報を生成する処理部である。具体的には、中間情報生成部17は、条件判定部16による判定結果をキーごとに組み合わせて中間情報DB12cに格納する。

30

## 【0036】

例えば、中間情報生成部17は、「キー」が「0001」のマスタデータについて「製品(True)、スタイル(False)、価格(True)」であることを条件判定部16から受信する。この場合、中間情報生成部17は、「キー、製品、スタイル、価格」として「0001、True、False、True」を生成して、中間情報DB12cに格納する。

## 【0037】

真偽判定表生成部18は、中間情報生成部17が生成した中間情報から真偽判定表を生成する処理部である。具体的には、真偽判定表生成部18は、中間情報DB12cから中間データを読み込み、抽出条件式DB12bから抽出条件式を読み込み、各中間データが条件を満たすか否かを判定し、その結果を真偽判定表13aとして第2記憶部13に格納する。

40

## 【0038】

例えば、真偽判定表生成部18は、中間データとして「キー(0001)、製品(True)、スタイル(False)、価格(True)」を中間情報DB12cから読み込む。また、真偽判定表生成部18は、図4示す抽出条件式「製品=PC AND (スタイル=ノート) or (価格=100000未満)」を抽出条件式DB12bから読み込む。

## 【0039】

50

そして、真偽判定表生成部18は、キーが「0001」の中間データについて、抽出条件に基づいて特定される「製品がTrueかつ（スタイルがTrue）または（価格がTrue）」を満たすか否かを判定する。この場合、真偽判定表生成部18は、「キー（0001）、製品（True）、スタイル（False）、価格（True）」であることから、製品かつ価格がTrueであるので、抽出条件を満たすと判定する。この結果、真偽判定表生成部18は、「キー、判定結果」として「0001、1」を対応付けて、真偽判定表13aに追加する。

#### 【0040】

ここでは、一例として真偽判定表生成部18が真偽判定表13aを生成する例を説明するが、これに限定されるものではない。例えば、真偽判定表生成部18は、真偽判定表13aではなく、「キー」と「判定結果」と対応付けたKVS（Key Value Store）形式のデータベースを生成することもできる。

10

#### 【0041】

ストリームデータ処理部19は、ストリームデータを受信し、真偽判定表13aに基づいてデータの抽出を実行する処理部である。具体的には、ストリームデータ処理部19は、受信したストリームデータからキーを抽出する。そして、ストリームデータ処理部19は、抽出したキーに対応する「判定結果」を真偽判定表13aから検索する。その後、ストリームデータ処理部19は、検索した「判定結果」が「1」である場合には、抽出対象と特定する。そして、ストリームデータ処理部19は、当該ストリームデータのキーに対応するマスタデータをマスタ情報DB12aから抽出する。一方、ストリームデータ処理部19は、検索した「判定結果」が「0」である場合には、抽出対象外と特定し、マスタデータの抽出を抑制する。

20

#### 【0042】

ここで具体例を説明する。図7は、ストリームデータの例を示す図である。図7のストリームデータは、一例としてPOSデータを例示した。図7に示すように、ストリームデータ送信装置1が送信するストリームデータは、「時間、キー、名前」から構成されるコード化されたデータである。「時間」は、ストリームデータが送信された時間またはストリームデータが生成された時間である。「キー」は、具体的なデータを特定するコードであり、マスタデータの検索に使用される。「名前」は、ストリームデータすなわちPOSデータを生成したユーザを示す。図7の場合、「2012-08-07 10:30」に生成された「キー」が「0003」で「名前」が「TOKKYO」であるストリームデータを示す。

30

#### 【0043】

例えば、図7のストリームデータと図6の真偽判定表を用いて具体例を説明する。ストリームデータ処理部19は、受信したストリームデータ「2012-08-07 10:30、0003、TOKKYO」からキーとして「0003」を抽出する。そして、ストリームデータ処理部19は、抽出したキー「0003」を検索キーにして、図6の真偽判定表を検索し、判定結果として「1」を特定する。

#### 【0044】

この結果、ストリームデータ処理部19は、当該ストリームデータを抽出対象のデータと判定する。そして、ストリームデータ処理部19は、ストリームデータ内のキー「0003」を検索キーにしてマスタ情報DB12aを検索して、「キー（0003）、製品（PC）、スタイル（Note）、価格（60000）」を抽出する。

40

#### 【0045】

##### [真偽判定表の生成例]

図8は、真偽判定表の生成例1を説明する図であり、図9は、真偽判定表の生成例2を説明する図である。図8に示すように、条件判定部16は、マスタ情報DB12aのキーが「0001」のマスタデータについて、製品（PC）、スタイル（Desktop）、価格（60000）の各々が抽出条件を満たすか否かを判定する。そして、条件判定部16は、それぞれについて条件を満たすか否かを示す「True」か「False」を決定する。

50

## 【0046】

続いて、中間情報生成部17は、キー「0001」のマスタデータについて、条件判定部16による判定結果「製品(True)、スタイル(False)、価格(True)」を対応付けた中間データを生成して、中間情報DB12cに格納する。

## 【0047】

その後、真偽判定表生成部18は、キー「0001」のマスタデータの中間データ「製品(True)、スタイル(False)、価格(True)」と、抽出条件とから、当該マスタデータを抽出対象と特定する。すると、真偽判定表生成部18は、「キー(0001)」と「True」とを対応付けた真偽判定表13aを生成する。

## 【0048】

続いて、図9に示すように、条件判定部16は、マスタ情報DB12aのキーが「0002」のマスタデータについて、製品(PC)、スタイル(Desktop)、価格(100000)の各々が抽出条件を満たすか否かを判定する。そして、条件判定部16は、それについて条件を満たすか否かを示す「True」か「False」を決定する。

## 【0049】

続いて、中間情報生成部17は、キー「0002」のマスタデータについて、条件判定部16による判定結果「製品(True)、スタイル(False)、価格(False)」を対応付けた中間データを生成して、中間情報DB12cに格納する。

## 【0050】

その後、真偽判定表生成部18は、キー「0002」のマスタデータの中間データ「製品(True)、スタイル(False)、価格(False)」と、抽出条件とから、当該マスタデータを抽出対象外と特定する。すると、真偽判定表生成部18は、「キー(0002)」と「False」とを対応付けた真偽判定表13aを生成する。このようにして、データ抽出装置10は、ストリームデータ受信前に、マスタ情報のデータ抽出判定を行って、キー毎に真偽判定表13aを生成する。

## 【0051】

## [処理の流れ]

図10は、実施例1に係るデータ抽出処理の流れを示すフローチャートである。図10に示すように、データ抽出装置10の条件判定部16は、処理が開始されると、すなわち、マスタ情報のメモリ展開が指示されると(S101:Yes)、抽出条件式DB12bから抽出条件式を読み出す(S102)。

## 【0052】

続いて、条件判定部16は、マスタ情報DB12aから1つのレコードを読み出して、すなわち、1つのマスタデータを読み出して、各カラムのデータが抽出条件を満たすか否かを判定する(S103)。

## 【0053】

そして、中間情報生成部17は、条件判定部16の結果に基づいて、中間情報を生成して、中間情報DB12cに格納する(S104)。続いて、真偽判定表生成部18は、中間情報DB12cに記憶される中間データと、抽出条件式DB12bに記憶される抽出条件式とから、真偽判定表13aを生成する(S105)。

## 【0054】

その後、マスタ情報DB12aに未判定のマスタデータが存在する場合には(S106:Yes)、データ抽出装置10は、S103に戻って以降の処理を繰り返す。一方、マスタ情報DB12aに未判定のマスタデータが存在しない場合には(S106:No)、真偽判定表生成部18は、真偽判定表13aをメモリへ展開する(S107)。例えば、真偽判定表生成部18は、生成した真偽判定表13aを第2記憶部13に格納する。

## 【0055】

その後、ストリームデータ処理部19は、ストリームデータを受信すると(S108:Yes)、ストリームデータからキーを抽出する(S109)。続いて、ストリームデータ処理部19は、抽出したキーを用いて真偽判定表13aを検索する(S110)。

10

20

30

40

50

## 【0056】

そして、ストリームデータ処理部19は、判定結果に「1」が格納されている場合(S111: Yes)、マスタ情報DB12aから、ストリームデータのキーに対応するマスタデータを抽出する(S112)。すなわち、ストリームデータ処理部19は、検索した真偽判定表13aが「1」である場合、マスタデータを抽出する。

## 【0057】

その後、ストリームデータ処理部19は、S108に戻って以降の処理を繰り返す。また、ストリームデータ処理部19は、判定結果に「0」が格納されている場合、すなわち、検索した真偽判定表13aが「0」である場合(S111: No)、マスタデータの抽出を行わずに、S108に戻って以降の処理を繰り返す。

10

## 【0058】

このように、実施例1に係るデータ抽出装置10は、ストリームデータ受信前に、マスタ情報のデータ抽出判定を行って、キー毎に真偽判定結果を作成する。そして、データ抽出装置10は、ストリームデータ受信時に、当該データ内のキーと真偽判定結果とを照合する。したがって、データ抽出装置10は、受信したストリームデータのキーを用いて、真偽判定表13aを検索することで、データ抽出の対象か否かを特定できるので、データ抽出を高速化することができる。

## 【0059】

また、データ抽出装置10は、ストリームデータ到着後に1回のキーの照合を行うだけで、抽出処理結果を得ることができるので、リアルタイム性を求められるストリームデータ処理において、処理遅延のリスクを削減できる。

20

## 【0060】

また、データ抽出装置10は、マスタデータに対する抽出条件の複雑度に依存せず、抽出処理結果を得ることができる。また、データ抽出装置10は、マスタデータに対する抽出条件の数に依存せず、抽出処理結果を得ることができる。データ抽出装置10は、マスタデータに対する抽出条件が複数の場合でも、判定表を1つ保持し、使用メモリ量を抑制できる。データ抽出装置10は、ストリームデータ到着後に演算処理を実行しないので、プロセッサの負荷を抑えることができる。

## 【実施例2】

## 【0061】

30

次に、真偽判定表13aを更新する例を説明する。ここでは、マスタ情報12aに記憶されるマスタデータが更新された場合に、真偽判定表13aを更新する例と、抽出条件式が更新された場合に、真偽判定表13aを更新する例とを説明する。

## 【0062】

## (マスタ情報の更新例)

図11は、実施例2に係るマスタ情報の削除例を説明する図である。図11では、マスタデータの更新によって真偽判定表13aの判定結果を更新する例を説明する。ここでは、一例としてマスタデータの削除を説明する。

## 【0063】

図11に示すように、データ抽出装置10の真偽判定表生成部18は、マスタ情報DB12aに格納されるマスタデータのうちキーが「0003」のマスタデータが削除されたことを検出する。すると、真偽判定表生成部18は、真偽判定表13aにおいてキー「0003」に対応付けられる判定結果に「-1(NULL)」を格納する。ここで、「-1」は、マスタデータが存在しないことを示す情報の一例である。

40

## 【0064】

このようにすることで、マスタデータが削除された場合でも、データ抽出装置10は、真偽判定表13aを作成しなおさずにデータ抽出処理を実行できるので、データ処理の遅延等を抑制できる。また、どのマスタデータが削除されたかを簡単に識別できるので、真偽判定表13aやマスタデータ等のデータ管理を効率的に行うことができる。

## 【0065】

50

## (抽出条件の更新例)

図12は、実施例2に係る抽出条件の更新例を説明する図である。図12に示すように、データ抽出装置10は、抽出条件式「製品 = P C AND (スタイル = ノート) or (価格 = 100000未満)」に基づいてデータ抽出判定を行った途中経過として、図5に示した中間データを記憶する。

## 【0066】

このような状態において、抽出条件式「製品 = P C AND (スタイル = ノート) or (価格 = 100000未満)」が抽出条件式「製品 = P C AND (スタイル = ノート) or (価格 = 50000未満)」に更新されたとする。つまり、価格の条件が「価格 < 100000」から「価格 < 50000」に更新されたとする。

10

## 【0067】

この場合、条件判定部16は、マスタ情報DB12aに記憶される各マスタデータについて、更新された「価格」についてのみ抽出条件を満たすか否かの判定処理を再度実行する。そして、中間情報生成部17は、条件判定部16が判定した「価格」の判定結果のみを、中間情報DB12cの中間データに反映させる。図12では、キー「0001」に対応する価格が「True」から「False」に変更される。

## 【0068】

その後、真偽判定表生成部18は、更新された中間データと、更新された抽出条件式とを用いて、真偽判定表13aを更新する。このように、データ抽出装置10は、検索条件式が更新された場合でも、更新された箇所について抽出判定をし直して、中間データを更新する。この結果、データ抽出装置10は、検索条件式が更新された場合でも、真偽判定表13aを高速に変更することができるので、データ処理の遅延等を抑制できる。

20

## 【実施例3】

## 【0069】

次に、図13から図16を用いて、真偽判定表13aを更新する際のストリームデータ処理について説明する。

## 【0070】

## (処理手法1)

図13と図14は、実施例3に係る真偽判定表更新時のストリームデータ処理例1を説明する図である。図13に示すように、データ抽出装置10は、ストリームデータを受信し、真偽判定表13aを用いてデータ抽出を行っている状況で、マスタ情報DB12aに記憶されるキーが「0004」のマスタデータを更新する(S201)。

30

## 【0071】

すると、データ抽出装置10は、ストリームデータの受信をブロックする(S202)。ここでブロックとは、例えば、受信を抑制してストリームデータ送信装置1に再送を行わせることや、ストリームデータを受信してデータ抽出処理を抑止することなどである。つまり、図13の場合、データ抽出装置10は、ストリームデータAについては受信またはデータ抽出を抑制する。この間、データ抽出対象ではあるが、未処理であるストリームデータA、ストリームデータBが滞留する。

## 【0072】

40

続いて、図14に示すように、データ抽出装置10は、キーが「0004」のマスタデータに対応する真偽判定表13aの判定結果を「False」から「True」に更新する(S203)。その後、データ抽出装置10は、ストリームデータの受信ブロックを解除する(S204)。

## 【0073】

このように、データ抽出装置10は、真偽判定表13aを更新する間、ストリームデータを用いたデータ抽出を抑制できる。したがって、データ抽出装置10は、最新の真偽判定表13aを用いて、データ抽出判定処理を実行することができ、判定ミス等を回避することができる。

## 【0074】

50

## (処理手法2)

図15と図16は、実施例3に係る真偽判定表更新時のストリームデータ処理例2を説明する図である。図15に示すように、データ抽出装置10は、真偽判定表13aを記憶するとともに、真偽判定表13aと同じデータで構成される真偽判定表13a'を記憶する。通常、データ抽出装置10は、真偽判定表13aを用いて、データ抽出判定処理を実行する。

## 【0075】

このような状態において、データ抽出装置10は、マスタ情報DB12aに記憶されるキーが「0004」のマスタデータを更新する(S301)。続いて、データ抽出装置10は、キーが「0004」のマスタデータに対応する真偽判定表13a'の判定結果を「False」から「True」に更新する(S302)。

## 【0076】

その後、図16に示すように、データ抽出装置10は、真偽判定表13a'の更新を完了すると、次から受信したストリームデータについては、真偽判定表13a'を用いて、データ抽出判定処理を実行する(S303)。例えば、データ抽出装置10は、真偽判定表を参照するポインタを切替えて、ストリームデータの入口をスイッチさせる。

## 【0077】

その後、データ抽出装置10は、真偽判定表13aについても同様に更新する。このように、データ抽出装置10は、真偽判定表の更新中も、受信をロックせずにストリームデータを処理することができる。この結果、データ抽出の高速化を維持できる。

## 【実施例4】

## 【0078】

さて、これまで本発明の実施例について説明したが、本発明は上述した実施例以外にも、種々の異なる形態にて実施されてよいものである。そこで、以下に異なる実施例を説明する。

## 【0079】

## (複数の抽出条件式)

例えば、データ抽出装置10は、複数の抽出条件式が抽出条件式DB12bに格納されている場合、条件式ごとに真偽判定表を生成することができる。図17は、真偽判定表の別例を説明する図である。

## 【0080】

図17に示すように、データ抽出装置10は、キー各々について、各条件式の判定結果を対応付けた真偽判定表を生成して保持する。そして、データ抽出装置10は、ストリームデータの種別、ストリームデータに含まれる識別子、管理者による事前指示等によって、参照する判定結果を動的に変更する。

## 【0081】

この結果、データ抽出装置10は、各条件式に真偽判定表を作成しないので、メモリ容量の削減にもつながり、1つのストリームデータで複数のデータ抽出を実行することができる、処理の高速化が図れる。

## 【0082】

## (中間情報)

実施例1等では、データ抽出装置10が中間情報を生成する例を説明したが、これに限定されるものではなく、中間情報を生成することなく、マスタ情報から真偽判定表を生成することもできる。

## 【0083】

## (真偽判定表)

実施例1等では、データ抽出装置10が、真偽判定表に、判定結果として「1(True)」、「0(False)」、「-1(NULL)」のいずれかを対応付ける例を説明したがこれに限定されるものではない。例えば、データ抽出装置10が、判定結果が「1(True)」であるデータだけを、真偽判定表に登録してもよい。こうすることで、デ

10

20

30

40

50

ータ抽出装置 10 は、データ抽出を高速化するとともに、メモリ容量の浪費を削減できる。

【 0 0 8 4 】

(システム)

また、本実施例において説明した各処理のうち、自動的におこなわれるものとして説明した処理の全部または一部を手動的におこなうこともできる。あるいは、手動的におこなわれるものとして説明した処理の全部または一部を公知の方法で自動的におこなうこともできる。この他、上記文書中や図面中で示した処理手順、制御手順、具体的な名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。

10

【 0 0 8 5 】

また、図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散や統合の具体的な形態は図示のものに限られない。つまり、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。さらに、各装置にて行なわれる各処理機能は、その全部または任意の一部が、CPU および当該 CPU にて解析実行されるプログラムにて実現され、あるいは、ワイヤードロジックによるハードウェアとして実現され得る。

【 0 0 8 6 】

(ハードウェア)

20

図 18 は、ハードウェア構成例を示す図である。図 18 に示すように、データ抽出装置 10 は、CPU (Central Processing Unit) 101、メモリ 102、HDD (Hard Disk Drive) 103、通信インターフェース 104、入力装置 105、表示装置 106 を有する。また、図 18 に示した各部は、バス等で相互に接続される。

【 0 0 8 7 】

HDD 103 は、図 2 に示した機能を動作させるプログラムやテーブルを記憶する。通信インターフェース 104 は、ネットワークインターフェースカードや無線インターフェースなどである。入力装置 105 は、例えばキーボードなどであり、表示装置 106 は、例えばタッチパネルやディスプレイなど、各種情報を表示する表示装置である。

【 0 0 8 8 】

30

CPU 101 は、図 2 に示した各処理部と同様の処理を実行するプログラムを HDD 103 等から読み出してメモリ 102 に展開することで、図 2 等で説明した各機能を実行するプロセスを動作させる。すなわち、このプロセスは、データ抽出装置 10 が有する各処理部と同様の機能を実行する。具体的には、CPU 101 は、条件判定部 16、中間情報生成部 17、真偽判定表生成部 18、ストリームデータ処理部 19 等と同様の機能を有するプログラムを HDD 103 等から読み出す。そして、CPU 101 は、中間情報生成部 17、真偽判定表生成部 18、ストリームデータ処理部 19 と同様の処理を実行するプロセスを実行する。

【 0 0 8 9 】

40

このようにデータ抽出装置 10 は、プログラムを読み出して実行することでデータ抽出方法を実行する情報処理装置として動作する。また、データ抽出装置 10 は、媒体読取装置によって記録媒体から上記プログラムを読み出し、読み出された上記プログラムを実行することで上記した実施例と同様の機能を実現することもできる。なお、この他の実施例でいうプログラムは、データ抽出装置 10 によって実行されることに限定されるものではない。例えば、他のコンピュータまたはサーバがプログラムを実行する場合や、これらが協働してプログラムを実行するような場合にも、本発明を同様に適用することができる。

【 0 0 9 0 】

以上の各実施例を含む実施形態に関し、さらに以下の付記を開示する。

【 0 0 9 1 】

(付記 1) コンピュータに、

50

キー情報とデータ情報を対応付けた各第1のデータにおける前記データ情報が、抽出条件を満たすか否かを判定し、

判定した前記データ情報に対応付けられる前記キー情報と判定結果とを対応付けた第2のデータを生成し、

受信データに含まれるキー情報を用いて、各第2のデータから前記判定結果を特定する処理を実行させるデータ抽出プログラム。

#### 【0092】

(付記2) 前記第2のデータを生成する処理は、前記各第1のデータのうち前記データ情報が前記抽出条件を満たすと判定された前記第1のデータのキー情報と判定結果とを対応付けた前記第2のデータを生成することを特徴とする付記1に記載のデータ抽出プログラム。

10

#### 【0093】

(付記3) 前記第2のデータを生成する処理は、前記第1のデータのデータ情報が更新された場合に、更新された前記第1のデータのキー情報に対応する判定結果を、更新後の前記データ情報に対する判定結果で更新し、前記第1のデータが削除された場合に、削除された前記第1のデータのキー情報に対応する判定結果を、削除されたことを示す削除情報で更新することを特徴とする付記1または2に記載のデータ抽出プログラム。

#### 【0094】

(付記4) 前記判定する処理が、前記データ情報が前記抽出条件を満たすか否かを判定するのに際して、前記データ情報を構成する各構成データについて、前記抽出条件を構成する各条件のうち当該構成データに対応する条件を満たすか否かを判定した第3のデータを生成し、

20

前記抽出条件が更新された場合に、更新された前記条件に対応する各構成データが当該条件を満たすか否かを判定して前記第3のデータを更新し、

前記更新した前記第3のデータと前記抽出条件とに基づいて、前記第2のデータの判定結果を更新する処理をさらにコンピュータに実行させることを特徴とする付記1から3のいずれか一つに記載のデータ抽出プログラム。

#### 【0095】

(付記5) コンピュータが、

キー情報とデータ情報を対応付けた各第1のデータにおける前記データ情報が、抽出条件を満たすか否かを判定し、

30

判定した前記データ情報に対応付けられる前記キー情報と判定結果とを対応付けた第2のデータを生成し、

受信データに含まれるキー情報を用いて、各第2のデータから前記判定結果を特定する処理を含むことを特徴とするデータ抽出方法。

#### 【0096】

(付記6) キー情報とデータ情報を対応付けた各第1のデータにおける前記データ情報が、抽出条件を満たすか否かを判定する判定部と、

前記判定部によって判定された前記データ情報に対応付けられる前記キー情報と判定結果とを対応付けた第2のデータを生成する生成部と、

40

受信データに含まれるキー情報を用いて、前記生成部によって生成された各第2のデータから前記判定結果を特定する特定部と

を有することを特徴とするデータ抽出装置。

#### 【0097】

(付記7) メモリと

前記メモリに接続されるプロセッサと、を有し、

前記プロセッサは、

キー情報とデータ情報を対応付けた各第1のデータにおける前記データ情報が、抽出条件を満たすか否かを判定し、

判定した前記データ情報に対応付けられる前記キー情報と判定結果とを対応付けた第2

50

のデータを生成し、

受信データに含まれるキー情報を用いて、各第2のデータから前記判定結果を特定する処理を実行するデータ抽出装置。

【0098】

(付記8) キー情報をデータ情報を対応付けた各第1のデータにおける前記データ情報が、抽出条件を満たすか否かを判定し、

判定した前記データ情報に対応付けられる前記キー情報を判定結果と対応付けた第2のデータを生成し、

受信データに含まれるキー情報を用いて、各第2のデータから前記判定結果を特定する処理をコンピュータに実行させるデータ抽出プログラムを記憶する、コンピュータ読み取り可能な記憶媒体。 10

【符号の説明】

【0099】

- 1 ストリームデータ送信装置
- 2 ネットワーク
- 10 データ抽出装置
- 11 通信部
- 12 第1記憶部
- 12 a マスター情報 D B
- 12 b 抽出条件式 D B
- 12 c 中間情報 D B
- 13 第2記憶部
- 13 a 真偽判定表
- 15 制御部
- 16 条件判定部
- 17 中間情報生成部
- 18 真偽判定表生成部
- 19 ストリームデータ処理部

10

20

【図1】

実施例1に係るシステムの全体構成例を示す図



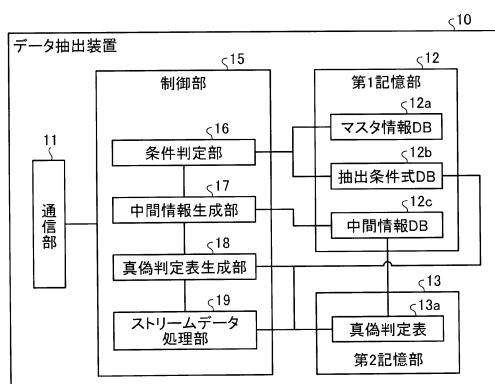
【図3】

マスタ情報DBに記憶される情報の例を示す図

キー	製品	スタイル	価格
0001	PC	Desktop	60000
0002	PC	Desktop	100000
0003	PC	Note	150000
0004	PC	Tablet	1000
0005	PC	Tablet	100000
...	...	...	...

【図2】

実施例1に係るデータ抽出装置の機能構成を示す機能ブロック図



【図4】

マスタ情報DBに記憶される情報の例を示す図

```

lookup("Master", Stream.$KEY == Master.$KEY, $製品) == "PC"
AND
lookup("Master", Stream.$KEY == Master.$KEY, $スタイル) == "Note"
OR
lookup("Master", Stream.$KEY == Master.$KEY, $価格) < 100000
  
```

【図5】

中間情報DBに記憶される情報の例を示す図

キー	製品	スタイル	価格
0001	True	False	True
0002	True	False	False
0003	True	True	False
0004	True	False	True
0005	True	False	False
...	...	...	...

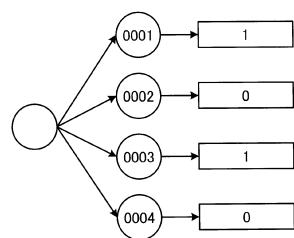
【図7】

ストリームデータの例を示す図

時間	キー	名前
2012-08-07 10:30	0003	TOKKYO

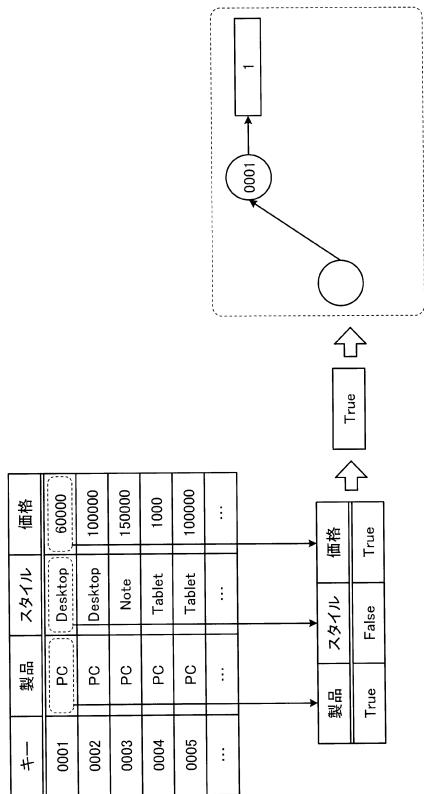
【図6】

真偽判定表の例を示す図



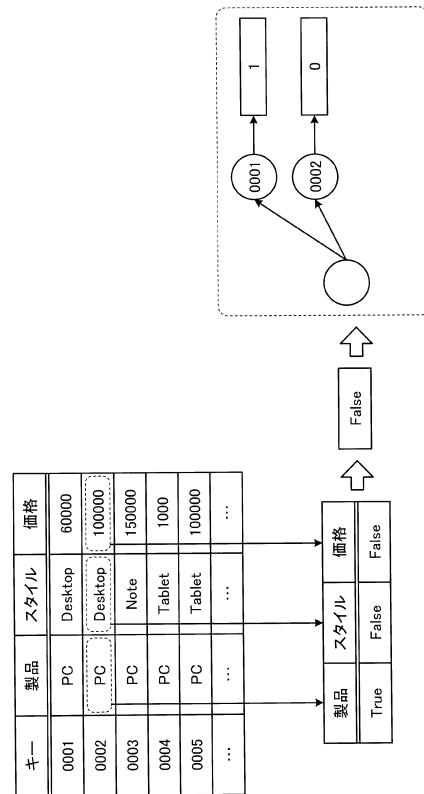
【図 8】

真偽判定表の生成例1を説明する図



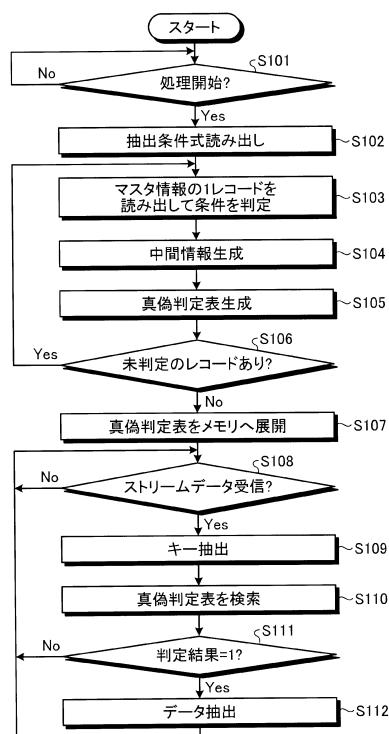
【図 9】

真偽判定表の生成例2を説明する図



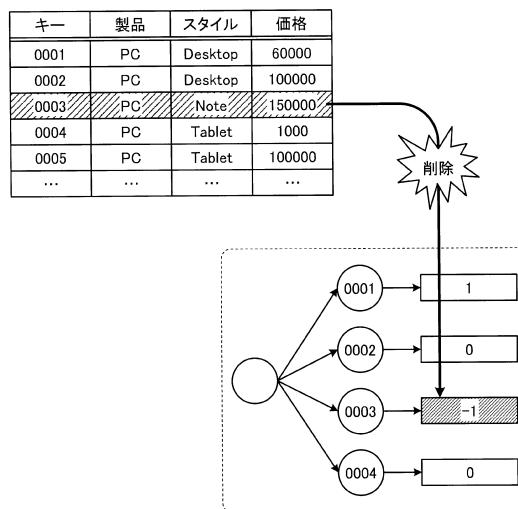
【図 10】

実施例1に係るデータ抽出処理の流れを示すフローチャート

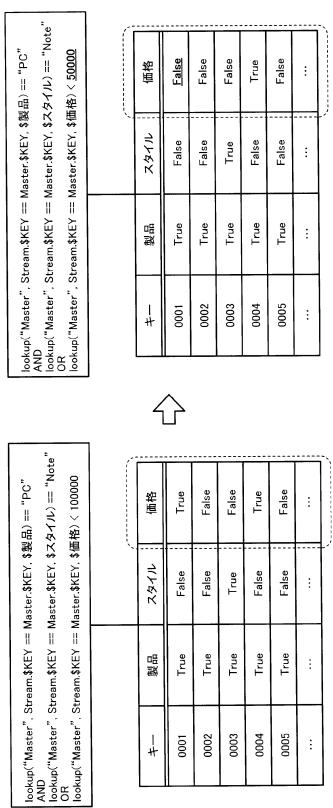


【図 11】

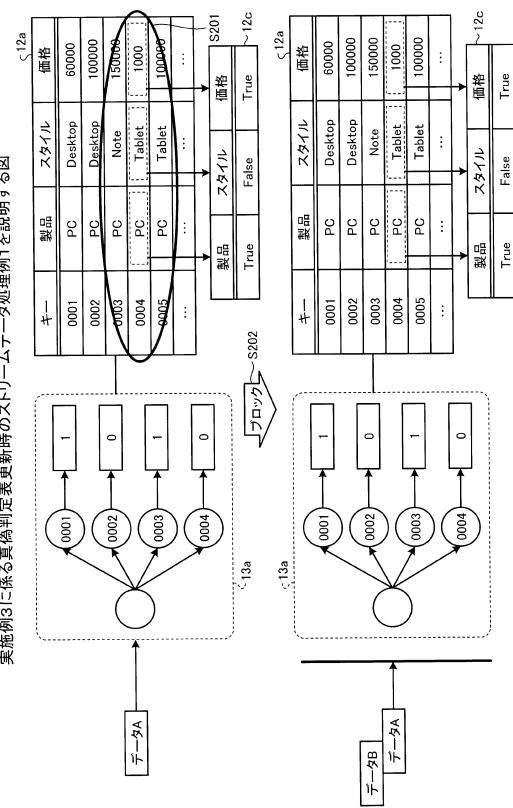
実施例2に係るマスタ情報の削除例を説明する図



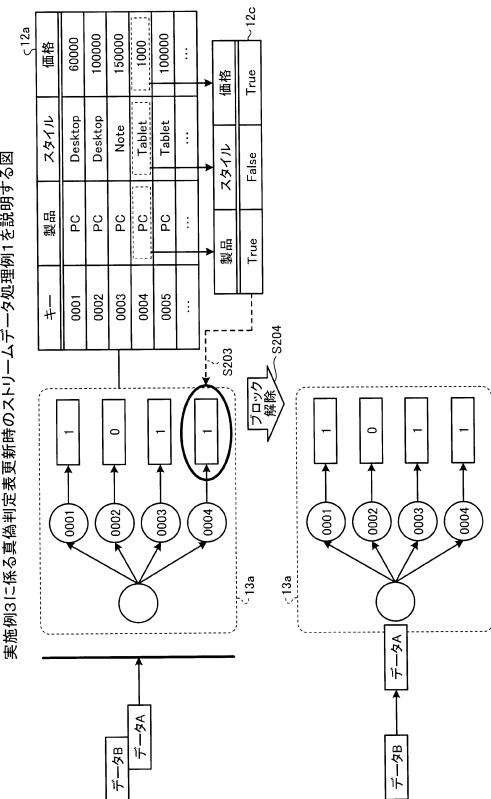
【図12】



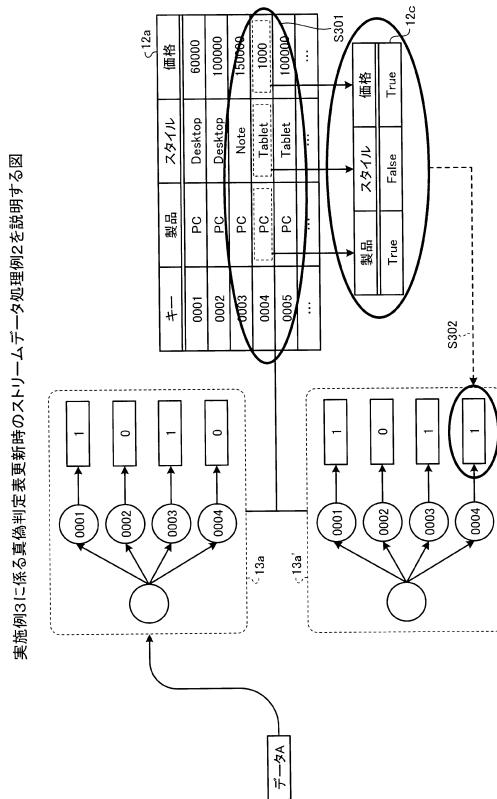
【図13】



【図14】

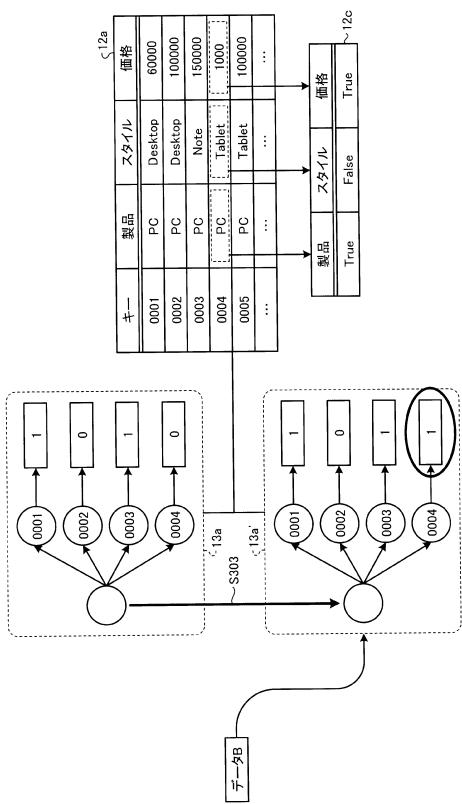


【図15】



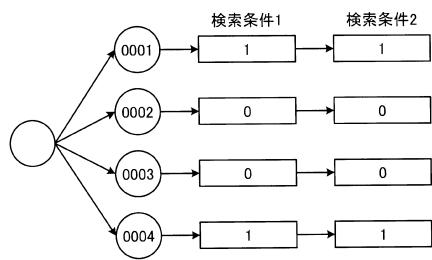
【図16】

実施例3に係る真偽判定表更新時のストリームデータ処理例を説明する図



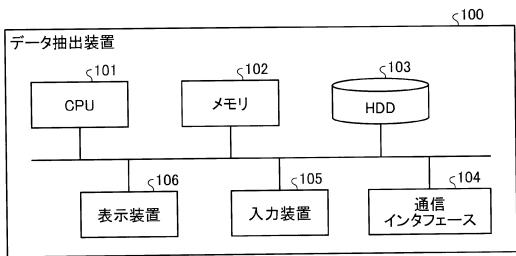
【図17】

真偽判定表の別例を説明する図



【図18】

ハードウェア構成例を示す図



---

フロントページの続き

(72)発明者 安岡 義弘  
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

(72)発明者 山田 樹一  
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

(72)発明者 佐々木 若菜  
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

審査官 田中 秀樹

(56)参考文献 國際公開第2012/164738 (WO, A1)

特開2006-215998 (JP, A)

國際公開第2010/095459 (WO, A1)

特開平05-324731 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G 06 F 17/30

G 06 F 12/00