

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7366763号  
(P7366763)

(45)発行日 令和5年10月23日(2023.10.23)

(24)登録日 令和5年10月13日(2023.10.13)

(51)国際特許分類

F I

G 0 6 F 16/215(2019.01)

G 0 6 F 16/215

請求項の数 8 (全18頁)

(21)出願番号	特願2020-5914(P2020-5914)	(73)特許権者	000005108
(22)出願日	令和2年1月17日(2020.1.17)		株式会社日立製作所
(65)公開番号	特開2021-114077(P2021-114077 A)	(74)代理人	110001689
(43)公開日	令和3年8月5日(2021.8.5)		青稜弁理士法人
審査請求日	令和4年6月9日(2022.6.9)	(72)発明者	塚野 夏樹
			東京都千代田区丸の内一丁目 6 番 6 号
		(72)発明者	株式会社日立製作所内
		(72)発明者	山本 暖
			東京都千代田区丸の内一丁目 6 番 6 号
		(72)発明者	株式会社日立製作所内
			川村 俊二
			東京都千代田区丸の内一丁目 6 番 6 号
			株式会社日立製作所内
		審査官	松尾 真人

最終頁に続く

(54)【発明の名称】 データ信頼度算出装置、データ信頼度算出方法およびデータ信頼度算出プログラム

(57)【特許請求の範囲】

【請求項 1】

データを利用するにあたってのデータ信頼度を算出するデータ信頼度算出装置であって、データのデータ利用者ごとのデータ利用者スコアと、データのデータ提供者ごとのデータ提供者スコアと、データごとのデータ信頼度とを保持し、

あるデータのデータ信頼度を算出する際に、そのデータを利用したデータ利用者のデータ利用者スコアと、そのデータのデータ提供者のデータ提供者スコアと、そのデータの元データのデータ信頼度に基づいて、そのデータのデータ信頼度を算出することを特徴とするデータ信頼度算出装置。

【請求項 2】

あるデータのデータ信頼度を算出する際に、そのデータを利用しているデータ利用者のデータ利用者スコアの総和と、そのデータのデータ提供者のデータ提供者スコアと、元データの信頼度の相加平均とを足し合わせた値を、そのデータのデータ信頼度として算出とすることを特徴とする請求項 1 記載のデータ信頼度算出装置。

【請求項 3】

元データのデータ I D と、加工後データのデータ I D と、データ加工の種別とを含むデータの来歴情報を保持し、

前記データ加工の種別ごと定まる係数を、各々の元データの信頼度に乗じて、元データの信頼度の相加平均をとることを特徴とする請求項 2 記載のデータ信頼度算出装置。

**【請求項 4】**

前記データ提供者スコアは、データ提供者が作成した全データのデータ信頼度の相加平均とし、

前記データ利用者スコアは、利用している全データのデータ信頼度の相加平均とし、データ信頼度として、データごとに初期値を与え、

与えられたデータのデータ信頼度からデータ利用者スコアを求める処理と、

与えられたデータのデータ信頼度からデータ提供者スコアを求める処理と、

前記データ利用者スコアと前記データ提供者スコアと前記元データの信頼度とに基づいて、データ信頼度を求める処理とを、

前記データ利用者スコアと前記データ提供者スコアと前記データ信頼度の全ての変化が一定の閾値に収まるまで、繰り返し行なうことを特徴とする請求項 2 記載のデータ信頼度算出装置。

10

**【請求項 5】**

さらに、データ利用者の所属情報、または、データ提供者の所属情報を保持し、

前記データ利用者の所属情報に基づいて、前記データ利用者スコアを、データ提供者の所属情報に基づいて、前記データ提供者スコアをそれぞれ定めることを特徴とする請求項 1 記載のデータ信頼度算出装置。

**【請求項 6】**

さらに、データが利用された日付情報、または、データが提供された日付情報を保持し、

前記データが利用された日付情報、または、データが提供された日付情報に基づいて、データ信頼度を定めることを特徴とする請求項 1 記載のデータ信頼度算出装置。

20

**【請求項 7】**

データ信頼度算出装置により、データを利用するにあたってのデータ信頼度を算出するデータ信頼度算出方法であって、

前記データ信頼度算出装置は、

データのデータ利用者ごとのデータ利用者スコアと、

データのデータ提供者ごとのデータ提供者スコアと、

データごとのデータ信頼度とを保持し、

あるデータのデータ信頼度を算出する際に、そのデータを利用しているデータ利用者のデータ利用者スコアの総和と、そのデータのデータ提供者のデータ提供者スコアと、元データの信頼度の相加平均とを足し合わせた値を、そのデータのデータ信頼度として算出するステップを有することを特徴とするデータ信頼度算出方法。

30

**【請求項 8】**

データ信頼度算出装置により実行され、データを利用するにあたってのデータ信頼度を算出するデータ信頼度算出プログラムであって、

前記データ信頼度算出装置は、

データのデータ利用者ごとのデータ利用者スコアと、

データのデータ提供者ごとのデータ提供者スコアと、

データごとのデータ信頼度とを保持し、

あるデータのデータ信頼度を算出する際に、そのデータを利用しているデータ利用者のデータ利用者スコアの総和と、そのデータのデータ提供者のデータ提供者スコアと、元データの信頼度の相加平均とを足し合わせた値を、そのデータのデータ信頼度として算出するステップを実行することを特徴とするデータ信頼度算出プログラム。

40

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は、データ信頼度算出装置、データ信頼度算出方法およびデータ信頼度算出プログラムに係り、特に、データを外部から取り込むデータ連携システムに関して、ユーザにデータ信頼性を即時に提示し、データの流通を促進するのに好適なデータ信頼度算出装置、データ信頼度算出方法およびデータ信頼度算出プログラムに関する。

50

**【背景技術】****【0002】**

近年、オープンデータやビッグデータなどのデータ活用に関する期待の高まりに伴い、データの流通、取引、連携に関する枠組みや情報処理システムによりデータを取り扱う技術の整備が進んでいる。また、インターネットの急速な普及に伴って公開されるデータの種類や量も年々増加している。

**【0003】**

データ流通においては、データを提供するデータ提供者、提供されるデータを利用するデータ利用者と、それらのデータ利用者とデータ提供者のやり取りを支援するデータ連携システムからなる仕組みが考えられている。

10

**【0004】**

このようなデータ連携システムの仕組みにおいて、データ利用者は、データを利用する前にデータを選定する。このとき、開発しようとしているアプリケーション等の要件に合致するかなどの基準に基づいて、データ提供者が公開しているデータ項目やデータ来歴などの品質について確認する。

**【0005】**

情報処理システムにより、データ選定を支援する従来技術としては、例えば、特許文献1がある。この特許文献1によると、データフロ制御装置が、デバイスが提供するデータの来歴を示したデバイス側メタデータと、アプリケーションが要求するデータの来歴を示したアプリ側メタデータをマッチングすることによって、複数のデバイスのなかからアプリケーションの要求する仕様としてデータを提供した来歴を有するデータを提供可能なデバイスを抽出し、これにより利用者が要求するデータの精度やデータ品質を保証することができるとしている。

20

**【先行技術文献】****【特許文献】****【0006】**

【文献】特開2017-111501号公報

**【発明の概要】****【発明が解決しようとする課題】****【0007】**

30

特許文献1に記載した従来技術は、IoT (Internet of Things) などの潮流を背景に、複数のセンサから出力されるセンシングデータを、アプリケーション側で最適なデータを選出するという点に着目している。

**【0008】**

しかしながら、特許文献1では、データ利用者は、アプリ側メタデータとして、データ項目だけでなく、アプリケーションが要求するデータの来歴も含めて設定する必要があり、これを実行しようとする、データの統計手法や加工手法、データのドメイン知識などの専門知識が必要になり、容易にデータを探索できない可能性がある。

**【0009】**

また、特許文献1は、データ利用者があるデータについてどの程度の信頼度をもって、データを利用してよいかという定量的な指標を示すことについて開示されていない。

40

**【0010】**

本発明の目的は、専門知識が少なくデータ来歴を記入できない、または提供データのデータ来歴の理解が難しいデータ利用者においても、使用しようとするデータがどの程度信頼できるかという定量的な指標を提示することできるデータ信頼度算出装置、データ信頼度算出方法およびデータ信頼度算出プログラムを提供することにある。

**【課題を解決するための手段】****【0011】**

本発明のデータ信頼度算出装置の構成は、好ましくは、データを利用するにあたってのデータ信頼度を算出するデータ信頼度算出装置であって、データのデータ利用者ごとのデ

50

ータ利用者スコアと、データのデータ提供者ごとのデータ提供者スコアと、データごとのデータ信頼度とを保持し、あるデータのデータ信頼度を算出する際に、そのデータを利用したデータ利用者のデータ利用者スコアと、そのデータのデータ提供者のデータ提供者スコアと、そのデータの元データのデータ信頼度に基づいて、そのデータのデータ信頼度を算出するようにしたものである。

【発明の効果】

【0012】

本発明によれば、専門知識が少なくデータ来歴を記入できない、または提供データのデータ来歴の理解が難しいデータ利用者においても、使用しようとするデータがどの程度信頼できるかという定量的な指標を提示できるデータ信頼度算出装置、データ信頼度算出方法およびデータ信頼度算出プログラムを提供することができる。

10

【図面の簡単な説明】

【0013】

【図1】データ信頼度算出システムの全体構成図である。

【図2】データ信頼度算出装置の機能構成図である。

【図3】データ連携装置の機能構成図である。

【図4】データ信頼度算出装置のハードウェア・ソフトウェア構成図である。

【図5】データ連携装置のハードウェア・ソフトウェア構成図である。

【図6】データカタログの一例を示す図である。

【図7】来歴者情報テーブルの一例を示す図である。

20

【図8】データ連携記録テーブルの一例を示す図である。

【図9】利用者情報テーブルの一例を示す図である。

【図10】データ信頼算出システムがデータ信頼度を算出し、表示する一連の処理を示すシーケンスである。

【図11】データ信頼度算出の処理を示すフローチャートである。

【図12A】データ信頼度算出処理の際に算出のデータ関連モデルを示すグラフである（その一）。

【図12B】データ信頼度算出処理の際に算出のデータ関連モデルを示すグラフである（その二）。

【発明を実施するための形態】

30

【0014】

本実施形態では、データの利用履歴やデータの来歴を利用してデータの信頼度を算出し、データ利用者に提示するシステムを説明する。

まず、図1ないし図5を用いてデータ信頼度算出システムの構成を説明する。

【0015】

データ信頼度算出システムは、図1に示されるように、インターネット等のネットワーク105に接続された、複数台の端末装置101と、データ信頼度算出装置100と、データ連携装置103と、複数台のデータ提供装置104から構成されるサービスシステムである。

【0016】

40

端末装置101（図では、端末装置101<sub>1</sub>...<sub>n</sub>と表記）は、データ利用者が、使用可能なデータを探索したり、他のサーバ装置のデータをアプリケーションソフトウェアで利用するための装置である。端末装置101では、データ利用者が、端末装置101にインストールされたWebブラウザ等を用いて使用可能なデータを探索したり、データ連携装置103または、データ提供装置104から取得したデータを、クライアントサイドまたはサーバサイドにおけるアプリケーションソフトウェアで利用することができる。端末装置101は、一般的なパーソナルコンピュータなどの情報処理装置により実現することができる。

【0017】

データ信頼度算出装置100は、データ連携装置103からの要請に基づいてデータの

50

信頼度を算出する装置である。

【0018】

データ連携装置103は、データ提供者とデータ利用者の連携を支援する装置である。

【0019】

データ提供装置104（図1では、データ提供装置104<sub>1</sub>...<sub>m</sub>と表記）は、データ利用者の端末装置101からの要求に応じてデータを提供する装置である。

【0020】

データ信頼度算出装置100、データ連携装置103、データ提供装置104は、共に、サーバ装置などの一般的な情報処理装置により実現することができる。

【0021】

また、端末装置101、データ信頼度算出装置100、データ連携装置103、データ提供装置104は、複数の装置の機能を一つのサーバ上に構築することも実現可能であり、また、それぞれの装置が保有する機能の一部を他の装置が担うことも可能である。例えば、データ信頼度算出装置100と、データ連携装置103は同一のサーバ上に構築してもよく、データ連携装置103のデータ連携機能部307のデータの送受信に関わる機能を、端末装置101およびデータ提供装置104に実現することもできる。

【0022】

次に、図2を用いてデータ信頼度算出装置の機能構成について説明する。

データ信頼度算出装置100は、図2に示されるように、機能構成として、データ処理部201、記憶部202、通信部203、入出力部204からなる。

【0023】

データ処理部201は、データ信頼度算出装置100で取扱うデータの演算や各種処理を行う機能部である。データ処理部201は、サブコンポーネントして、連携結果取得部205、データカタログ取得部206、利用者情報取得部207、データ信頼度算出部208、データ信頼度登録部209を備える。

【0024】

連携結果取得部205は、データ連携装置103から、データの連携結果を取得する機能部である。データカタログ取得部206は、データカタログ（後述）を取得する機能部である。利用者情報取得部207は、データ提供者およびデータ利用者の情報を取得する機能部である。データ信頼度算出部208は、外部から取得したデータに基づいてデータ信頼度を算出する機能部である。データ信頼度登録部209は、算出したデータ信頼度をデータ連携装置103のデータカタログに登録する機能部である。

【0025】

記憶部202は、データ信頼度算出装置100で取扱うデータを記憶する機能部である。記憶部202は、データカタログ401、データ来歴情報テーブル402、データ連携記録テーブル403、利用者情報テーブル404を保持する。なお、各種テーブルの詳細は、後述する。

【0026】

通信部203は、ネットワーク105を介して、データ連携装置103と通信を行う機能部である。入出力部204は、入出力装置により、管理者からデータやコマンドを外部から入力したり、管理者に情報の出力を行う機能部である。

【0027】

次に、図3を用いてデータ連携装置の機能構成について説明する。

データ連携装置103は、図2に示されるように、機能構成として、データ処理部301、記憶部302、通信部303、入出力部304からなる。

【0028】

データ処理部301は、サブコンポーネントとして、利用者管理部305、データカタログ管理部306、データ連携機能部307を備える。

【0029】

利用者管理部305は、データ提供者やデータ利用者などのアカウントの登録、更新、

10

20

30

40

50

削除したり、パスワードや電子証明書等の認証情報の管理を行う機能部である。データカタログ管理部 306 は、データの識別子、タイトル、作成者、来歴情報等の情報を管理する機能部である。データ連携機能部 307 は、データ利用者とデータ提供者とのデータのやり取りを支援する機能部である。

#### 【0030】

記憶部 302 は、データ連携装置 103 で取扱うデータを記憶する機能部である。記憶部 302 は、データカタログ 401、データ来歴情報テーブル 402、データ連携記録テーブル 403、利用者情報テーブル 404 を保持する。なお、各種テーブルの詳細は、後述する。

#### 【0031】

通信部 303 は、ネットワーク 105 を介して、端末装置 101、データ提供装置 104、データ信頼度算出装置 100 と通信を行う機能部である。入出力部 204 は、入出力装置により、管理者からデータやコマンドを外部から入力したり、管理者に情報の出力を行う機能部である。

#### 【0032】

データ利用者とデータ提供者とのデータのやり取りの際には、データ連携装置 103 を経由してデータのやり取りを行う場合と、データ連携装置 103 を経由せずに、データ利用者とデータ提供者とが直接データのやり取りを行う場合とが考えられる。いずれの場合も、データ連携の結果は、データ連携装置 103 に記録される。例えば、データ連携装置 103 を経由する場合、データ連携機能部 307 は、データ利用者のデータ要求に応じてデータ提供した結果を、データ連携記録テーブル 403 として、記憶部 302 に記録する。データ連携装置 103 を経由しない場合は、端末装置 101 にインストールされたアプリケーションまたはデータ提供装置 104 にインストールされたアプリケーションから、データ連携装置 103 に対して、データの連携結果が送信され、データ連携結果を受信したデータ連携装置 103 は、データ連携記録テーブル 403 に連携結果を登録する。

#### 【0033】

次に、図 4 を用いてデータ信頼度算出装置のハードウェア・ソフトウェア構成について説明する。

データ信頼度算出装置 100 のハードウェア構成としては、例えば、図 4 に示されるサーバ装置のような一般的な情報処理装置で実現される。

#### 【0034】

データ信頼度算出装置 100 は、CPU (Central Processing Unit) 502、主記憶装置 504、ネットワーク I/F (InterFace) 506、表示 I/F 508、入出力 I/F 510、補助記憶 I/F 512 がバスにより結合された形態になっている。

#### 【0035】

CPU 502 は、データ信頼度算出装置 100 の各部を制御し、主記憶装置 504 に必要なプログラムをロードして実行する。

#### 【0036】

主記憶装置 504 は、通常、RAM などの揮発メモリで構成され、CPU 502 が実行するプログラム、参照するデータが記憶される。

#### 【0037】

ネットワーク I/F 506 は、ネットワーク 105 と接続するためのインタフェースである。

#### 【0038】

表示 I/F 508 は、LCD (Liquid Crystal Display) などの表示装置 520 を接続するためのインタフェースである。

#### 【0039】

入出力 I/F 510 は、入出力装置を接続するためのインタフェースである。図 4 の例では、キーボード 530 とポインティングデバイスのマウス 532 が接続されている。

#### 【0040】

10

20

30

40

50

補助記憶 I / F 5 1 2 は、H D D ( Hard Disk Drive ) 5 5 0 や S S D ( Solid State Drive ) などの補助記憶装置を接続するためのインタフェースである。

【 0 0 4 1 】

H D D 5 5 0 は、大容量の記憶容量を有しており、本実施形態を実行するためのプログラムが格納されている。データ信頼度算出装置 1 0 0 の H D D 5 5 0 には、連携結果取得プログラム 5 6 0、データカタログ取得プログラム 5 6 1、利用者情報取得プログラム 5 6 2、データ信頼度算出プログラム 5 6 3、データ信頼度登録プログラム 5 6 4 がインストールされている。

【 0 0 4 2 】

連携結果取得プログラム 5 6 0、データカタログ取得プログラム 5 6 1、利用者情報取得プログラム 5 6 2、データ信頼度算出プログラム 5 6 3、データ信頼度登録プログラム 5 6 4 は、それぞれ、連携結果取得部 2 0 5、データカタログ取得部 2 0 6、利用者情報取得部 2 0 7、データ信頼度算出部 2 0 8、データ信頼度登録部 2 0 9 の各機能を実現するプログラムである。

10

【 0 0 4 3 】

また、データ信頼度算出装置 1 0 0 の H D D 5 5 0 には、データカタログ 4 0 1、データ来歴情報テーブル 4 0 2、データ連携記録テーブル 4 0 3、利用者情報テーブル 4 0 4 が格納されている。

【 0 0 4 4 】

次に、図 5 を用いてデータ信頼度算出装置のハードウェア・ソフトウェア構成について説明する。

20

データ連携装置 1 0 3 も、データ信頼度算出装置 1 0 0 と同様に、ハードウェア構成としては、例えば、図 5 に示されるサーバ装置のような一般的な情報処理装置で実現され、そのハードウェア構成も同様である。

【 0 0 4 5 】

データ連携装置 1 0 3 の H D D 6 5 0 には、利用者管理プログラム 6 6 0、データカタログ管理プログラム 6 6 1、データ連携機能プログラム 6 6 2 がインストールされている。

【 0 0 4 6 】

利用者管理プログラム 6 6 0、データカタログ管理プログラム 6 6 1、データ連携機能プログラム 6 6 2 は、それぞれ利用者管理部 3 0 5、データカタログ管理部 3 0 6、データ連携機能部 3 0 7 の各機能を実現するプログラムである。

30

【 0 0 4 7 】

また、データ連携装置 1 0 3 の H D D 6 5 0 には、データカタログ 4 0 1、データ来歴情報テーブル 4 0 2、データ連携記録テーブル 4 0 3、利用者情報テーブル 4 0 4 が格納されている。

【 0 0 4 8 】

次に、図 6 ないし図 9 を用いて本実施形態のデータ信頼算出システムで用いられるデータ構造について説明する。

【 0 0 4 9 】

データカタログ 4 0 1 は、データに関するデータに関する基本的な情報を保持するテーブルであり、図 6 示されるように、[ D a t a I D ] 4 0 1 a、[ T i t l e ] 4 0 1 b、[ T r u s t \_ S c o r e ] 4 0 1 c、[ S u b \_ S c o r e ] 4 0 1 d、[ P r o v i d e r ] 4 0 1 f の各項目を保持する。

40

【 0 0 5 0 】

[ D a t a I D ] 4 0 1 a は、データの一意的な識別子が格納される項目である。[ T i t l e ] 4 0 1 b は、データ提供者がデータにつけた名称が格納される項目である。例えば、[ D a t a I D ] 4 0 1 a に感応されるデータ ID として、D 0 0 1、D 0 0 2、D 0 0 3 などシリアル番号を使うことが考えられるが、[ T i t l e ] 4 0 1 b に格納されるデータにつけた名称としては、利用者がデータの中身を想像できるものが望ましく、例えば、図 6 に示されるように、W e a t h e r、T e m p e r a t u r e、T r a f f

50

i c j a mなどの値が一例として挙げられる。

【0051】

[ T r u s t \_ S c o r e ] 4 0 1 c には、データ信頼度算出装置により算出されたデータの信頼度であり定量的な値が格納される。データ信頼度の算出処理は、後に詳述する。[ S u b \_ S c o r e ] 4 0 1 d には、[ T r u s t \_ S c o r e ] 4 0 1 c に対しての補助的なデータ信頼度が格納される。例えば、[ S u b \_ S c o r e ] 4 0 1 d として、当該データのデータ提供者のスコア、データの来歴をもとにした前データのスコア、データの使用頻度等を設定することができる。また、[ S u b \_ S c o r e ] 4 0 1 d の補助的なデータ信頼度としては、W e b で情報公開を行ったときのそのデータに関する評価などを参酌して定めてもよい。

10

【0052】

[ P r o v i d e r ] 4 0 1 e には、当該データを提供しているデータ提供者の情報を保持する。例えば、データ提供者の U s e r I D が設定される。

【0053】

なお、データカタログ 4 0 1 の項目としては、上記に限定されず、データの作成者、データの更新頻度、データの精度、最終更新日、権利などの項目を有してもよく、それらについても、データ信頼度の算出に用いることができる。

【0054】

次に、データ来歴情報テーブル 4 0 2 は、データの来歴を保持するテーブルとあり、図 7 に示されるように、[ D a t a I D ] 4 0 2 a、[ I n p u t \_ D a t a \_ I n f o ] 4 0 2 b、[ R e l a t i o n ] 4 0 2 c の項目を保持する。

20

【0055】

[ D a t a I D ] 4 0 2 a は、データの一意的な識別子が格納される項目である。[ I n p u t \_ D a t a \_ I n f o ] 4 0 2 b は、該当するデータの加工前のデータあるいは由来となるデータとなるデータ ID が保持される。[ R e l a t i o n ] 4 0 2 c には、[ I n p u t \_ D a t a \_ I n f o ] 4 0 2 b のデータ ID の示すデータと、[ D a t a I D ] 4 0 2 a の示すデータとの関係が示される。

【0056】

データ来歴情報テーブル 4 0 2 に保持される来歴情報としては、データを作成する際に使用したデータの情報や、センサ情報、加工方法を管理することが考えられる。例えば、図 7 では D 0 0 4 で示されるデータは、作成する際に、D 0 0 3 で示されるデータが使用され、加工処理としては分析処理 1 ( A n l y s i s 1 ) の処理が行われたことが例示されている。また、D 0 0 5 で示されるデータは、作成する際に、D 0 0 3 で示されるデータが使用され、加工処理としては抜粋処理 ( E x t r a c t ) が行われたことが例示されている。

30

【0057】

次に、データ連携記録テーブル 4 0 3 は、データの提供・使用に関する情報を保持するテーブルであり、図 8 に示されるように、[ D a t a I D ] 4 0 3 a、[ U s e r I D ] 4 0 3 b、[ R e l a t i o n ] 4 0 3 d、[ D a t e ] 4 0 3 e の項目を保持する。

【0058】

[ D a t a I D ] 4 0 3 a は、データの一意的な識別子が格納される項目である。[ U s e r I D ] 4 0 3 b は、そのデータを提供または利用したユーザの一意的な識別子が格納される項目である。[ R e l a t i o n ] 4 0 3 d には、そのデータを提供か使用したかを示す情報が格納される。[ D a t e ] 4 0 3 e には、そのデータを提供または使用した日付を示す情報が格納される。

40

【0059】

これらの情報により、[ D a t a I D ] 4 0 3 a のデータ ID で示されるデータが、[ U s e r I D ] 4 0 3 のユーザ ID で示される利用者によって、提供された、または利用されたことが示される。例えば、図 8 では、D 0 0 1 で示されるデータが、P 0 0 1 で示されるユーザ ( 提供者 ) によって、2 0 1 9 - 9 - 1 に提供されたこと。および、D 0 0

50



1で示されるデータが、U002で示されるユーザ（利用者）によって、2019-9-4に利用されたことが例示されている。

#### 【0060】

次に、利用者情報テーブル404は、データ提供者およびデータ利用者に関する基本的な情報を保持するテーブルであり、図8に示されるように、[UserID]402a、[Name]402b、[Organization]402cの項目を保持する。[UserID]402aは、そのデータを利用したユーザの一意的な識別子が格納される項目である。[Name]402bは、そのデータを利用したユーザの名前が格納される。なお、[Name]402bには、利用者個人の名前が設定される場合、部署名等が設定される場合などが考えられる。[Organization]402cには、[Name]402bに格納された個人の属する機関の情報、また、[Name]402bに格納された部署名等が属する会社名などの情報が格納される。

10

#### 【0061】

次に、図10を用いて、データ信頼算出システムがデータ信頼度を算出し、表示する一連の処理について説明する。

この処理は大きく、三つのフェーズに分割することができる。具体的には、(I)データ提供者がデータカタログを登録するフェーズ、(II)データ利用者がデータを利用するフェーズ、(III)データ信頼度が算出され、表示されるフェーズである。

#### 【0062】

(I)データ提供者がデータカタログを登録するフェーズ

20

まず、データ利用者が、データ提供装置104とデータ連携装置103に利用者情報の登録を行う(S500a、S500b)。

#### 【0063】

次に、データ提供装置104は、データ連携装置103にデータカタログの情報および来歴情報を登録し(S501)、データ連携装置103は、データ提供者の情報とそれらを合わせて記録する。

#### 【0064】

次に、データ信頼度算出装置100は、データ連携装置103のデータカタログが登録されたタイミングまたは日次などの所定のタイミングでデータ連携装置103からデータカタログを取得する(S502)。

30

#### 【0065】

(II)データ利用者がデータを利用するフェーズ

次に、データ利用者は、自身が利用したいデータを保持するデータ提供者を特定し、データ信頼度を確認したうえで、データ提供装置104からデータを取得する(S503)。データを取得した端末装置101は、データの取得結果をデータ連携装置103に送信する(S504)。一方で、データを提供したデータ提供装置104もデータの提供結果をデータ連携装置103に送信する(S505)。データ連携装置103は、端末装置101と、データ提供装置104から受信したデータをもとに、データの連携結果を記録する(S506)。データ信頼度算出装置100は、データ連携結果が記録されたタイミングまたは定められた日時などの所定のタイミングで、データ連携装置103からデータ連携記録を取得する(S507)。

40

#### 【0066】

なお、データ利用者がデータを取得する場合に、端末装置101がデータ提供装置104から直接データを取得する以外にも、データ提供者がデータ提供装置104からデータ連携装置103にデータを送信し、データ連携装置103からデータを取得するような場合も考えられる。なお、端末装置101にデータ連携の結果を送信するアプリケーションがインストールされていない場合は、データ連携装置103は、データ提供装置104から受信した情報を用いてデータ連携記録を作成する。

#### 【0067】

(III)データ信頼度が算出され、表示されるフェーズ

50

次に、データ信頼度算出装置 100 は、取得済みのデータカタログと、データの連携結果とをもとに、データ信頼度を算出し (S508)、算出したデータ信頼度をデータ連携装置 103 に登録する (S509)。データ利用者は、端末装置 101 を用いてデータ連携装置 103 からデータカタログの情報を表示し、それぞれのデータのデータ信頼度を確認する。(S509)。

【0068】

次に、図 11 ないし図 12B を用いて、データ信頼度算出装置が、データ信頼度を算出する処理について説明する。

【0069】

データ信頼度算出の処理は、図 10 に示したように、データ信頼度算出装置 100 が、各種テーブルを参照して行う処理であり、図 10 の S508 に該当する処理である。

【0070】

本実施形態では、データ信頼度を算出するために、データ信頼度に関する以下の前提を設ける。

- 1) 多くの利用者に使用されているデータは信頼性が高い。
- 2) 信頼性が高いデータを作成したデータ提供者が提供する他のデータは、ある程度信頼性が高い。
- 3) 信頼性が高いデータから加工されたデータは、ある程度信頼性が高い。

【0071】

データ信頼度を算出する処理は、データ提供者およびデータ利用者のスコア算出にデータの信頼度を用い、データの信頼度の算出にデータ利用者およびデータ提供者のスコアを用いるため、データの相互参照が再帰的に行われる処理である。処理の終了の判定には、データ更新前とデータ更新後のデータ信頼度、および、それに関連するデータ提供者スコア、データ利用者スコアを比較し、一定の閾値以下になった場合に、それらの値が収束したと判定し、処理を終了する。

【0072】

まず、データ信頼度算出装置は、記憶部 202 に記憶されたデータを取得し (S601)、図 12A に示すような、データ関連グラフ 700 を作成する (S602)。具体的には、データカタログ 401 からデータの一覧と、利用者情報テーブル 404 から取得したデータ利用者およびデータ提供者一覧とから、データに対応するノード、データ利用者に対応するノード、データ提供者に対応するノードを配置する。図 12A の DXXX としたノード (D ノード) がデータに対応するノードであり、PXXX としたノードがデータ提供者に対応するノード (P ノード) であり、UXXX としたノード (U ノード) がデータ利用者に対応するノードである。各々のノードの外周に記載した数値は、それぞれ、データ信頼度、データ提供者スコア、データ利用者スコアである。また、以下ではノード DXXX のノードが表すデータを単に、「データ DXXX」のように表現する。

【0073】

次に、データ連携記録テーブル 403 の情報に基づいて、データを利用したデータ利用者と当該データとを全てエッジでつなぐ。例えば、データ D001 を使用しているデータ利用者 U001 をつなぐ。次に、データを提供したデータ提供者と当該データを全てエッジでつなぐ。例えば、データ D001 を提供したデータ提供者 P001 とを全てエッジでつなぐ。次に、データ来歴情報テーブル 402 の記録をもとに、あるデータの元データと当該データとを全てエッジでつなぐ。例えば、あるデータ D004 とその元となったデータ D003 をエッジでつなぐ。

【0074】

次に、全てのノードについて、S604 ~ S606 の処理を繰り返す (S603 ~ S607)。

ループでは、まず、作成したデータ関連グラフ 700 に基づきデータ利用者スコアを算出する (S604)。

【0075】

10

20

30

40

50

データ利用者スコアの例としては、利用している全データのデータ信頼度の相加平均とすることが考えられる。

【 0 0 7 6 】

次に、データ提供者スコアを算出する（ S 6 0 5 ）。データ提供者スコアの例としては、データ提供者が作成した全データのデータ信頼度の平均とすることが考えられる。

【 0 0 7 7 】

次に、データの信頼度を算出する（ S 6 0 6 ）。データの信頼度の例としては、当該データを利用しているデータ利用者のデータ利用者スコアの総和と、当該データのデータ提供者のデータ提供者スコアと、来歴から取得した元データの信頼度の相加平均とを足し合わせることを考えられる。

【 0 0 7 8 】

すなわち、データ d の信頼度は、以下の（式 1 ）で表される。

【 0 0 7 9 】

【数 1 】

$$T(d) = \sum_{i=1}^n Su_i(d) + Sp(d) + \sum_{i=1}^m \frac{a_i T(d_i)}{m} \quad \dots(\text{式}1)$$

【 0 0 8 0 】

ここで、 $Su_i(d)$ （ $i = 1, \dots, n$ ）は、データ d を利用しているデータ利用者のデータ利用者スコアであり、 $Sp(d)$  は、データ d の提供者のデータ提供者スコア、 $d_i$ （ $i = 1, \dots, m$ ）は、データ d の元データ、 $T(d_i)$  は、データ d の元データのデータ信頼度である。これは、そのデータを利用した者が多いデータには、データ信頼度が高い、また、信頼度が高い（データ提供者スコアが高い）データ提供者のデータには、そのデータ信頼度が高いとみなせるということを参酌したものである。また、元データのデータ信頼度が高いデータは、そのデータ信頼度が高いとみなせるということを参酌したものである。なお、元データが複数ある場合があること（データをマージして一つのデータとすることなど）を考慮して、ここでは、元データのデータ信頼度の相加平均をとっている。

【 0 0 8 1 】

このとき、来歴情報として加工履歴がない、初めて作られたデータの場合には、所定の初期値を設定する。また、加工方法が信頼度に影響すると考え、加工方法ごとに異なる係数  $a_i$ （ $0 < a_i < 1$ ）を、元データの信頼度に乗算する。例えば、データをマージした場合には、それぞれの係数を  $a_i = 0.9$  とし、データにほとんど手を加えず抜粋処理のみ場合には、その係数を  $a_i = 0.8$  とし、データを何らかの統計手法で変更した結果には、その係数を  $a_i = 0.3$  とするなどが考えられる。

【 0 0 8 2 】

次に、S 6 0 3 ないし S 6 0 7 で得られたそれぞれのデータ利用者スコア、データ提供者スコアおよびデータ信頼度について、データ利用者スコア、データ提供者スコアおよびデータ信頼度ごとに正規化する（ S 6 0 8 ）。

【 0 0 8 3 】

ここで、正規化とは、各々の足し合わせた数を 1 になるように、各々の値を配分することであり、以下の（式 2 ）～（式 4 ）で表される。

【 0 0 8 4 】

【数 2 】

10

20

30

40

50

$$T'(d) = \sum_{i: \text{全てのDノード}} \frac{T(d)}{T(d_i)} \quad \dots(\text{式2})$$

$$Su'(u) = \sum_{i: \text{全てのUノード}} \frac{Su(u)}{Su(u_i)} \quad \dots(\text{式3})$$

$$Sp'(p) = \sum_{i: \text{全てのPノード}} \frac{Sp(p)}{Sp(p_i)} \quad \dots(\text{式4})$$

10

20

## 【 0 0 8 5 】

ここで、 $T'(d)$  は、正規化後のデータ  $d$  のデータ信頼度、 $T(d)$  は、正規化前のデータ  $d$  のデータ信頼度であり、(式2)の分母の  $T(d_i)$  は、全てのDノードに対して和をとることを意味する。同様に、 $Su'(u)$  は、正規化後のデータ利用者  $u$  のデータ利用者スコア、 $Su(u)$  は、正規化前のデータ利用者  $u$  のデータ利用者スコアであり、(式3)の分母の  $Su(u_i)$  は、全てのUノードに対して和をとることを意味する。また同様に、 $Sp'(p)$  は、正規化後のデータ提供者  $p$  のデータ提供者スコア、 $Sp(p)$  は、正規化前のデータ提供者  $p$  のデータ提供者スコアであり、(式4)の分母の  $Sp(p_i)$  は、全てのPノードに対して和をとることを意味する。

30

## 【 0 0 8 6 】

次に、全てのデータ利用者スコア、データ提供者スコアおよびデータ信頼度について、更新前(前回の更新時の値)と更新後の差分を計算し、全ての差分が閾値未満であるとき(S609: YES)。データ利用者スコア、データ提供者スコアおよびデータ信頼度を、データカタログ401に記録して終了する(S610)。

## 【 0 0 8 7 】

なお、S609で一つのデータ利用者スコア、データ提供者スコアおよびデータ信頼度の差分でも閾値以上である場合には(S609: NO)、各々のノードにおけるデータ利用者スコア、データ提供者スコアおよびデータ信頼度を更新し(S611)、ステップ603に戻る。

40

## 【 0 0 8 8 】

図12Bの各ノードIDの外周に、全ての差分が閾値未満であったときのそれぞれのデータ利用者スコア、データ提供者スコアおよびデータ信頼度が記載されている。例えば、データD001のデータ信頼度は0.428と算出され、データD002のデータ信頼度は0.168となる。D002とD004は、共にデータ利用者がいないが、データD002は、P001を介してD001とつながっており、信頼度の高いデータD001を提供しているP001が同じく提供しているD002も信頼度を高くなっている。

## 【 0 0 8 9 】

なお、S606においてデータの信頼度を算出する際に、データ連携記録テーブル40

50

3を用い、データの使用日によって、ユーザスコアを増減することが考えられる。例えば、データの使用日が一年以上前の場合にデータ信頼度に加算するユーザのスコアを半減することなどが考えられる。

【0090】

また、S604、S604のデータ利用者スコアを算出する処理、データ提供者スコアを算出する処理において、データ利用者またはデータ提供者の所属を考慮することが考えられる。例えば、利用者情報テーブル404において、AliceはAAA Ltd.に所属していることに基づいて、AAA Ltd.がどの程度信頼できるかをデータ利用者スコアに反映するなどすることができる。データ提供者の所属する情報を示すテーブルは、図示しなかったが、利用者情報テーブル404と同様のデータ提供者情報テーブルを用意し、データ提供者ごとに所属を示せばよい。

10

【0091】

以上、本実施形態によれば、全てのデータ利用者は、データの利用頻度やデータとデータ提供者の関係、データとデータ利用者との関係を模したグラフに基づいて算出されたデータ信頼度によって、データがどの程度信頼できるかを簡易に判断することができる。

【符号の説明】

【0092】

100...データ信頼度算出装置、101...端末装置、103...データ連携装置、104...データ提供装置、105...ネットワーク、201...データ処理部、202...記憶部、203...通信部、204...入出力部、205...連携結果取得部、206...データカタログ取得部、207...利用者情報取得部、208...データ信頼度算出部、209...データ信頼度登録部、

20

301...データ処理部、302...記憶部、303...通信部、304...入出力部、305...利用者管理部、306...データカタログ管理部、307...データ連携機能部、401...データカタログ、402...データ来歴情報テーブル、403...データ連携記録テーブル、404...利用者情報テーブル、700...データ関連グラフ

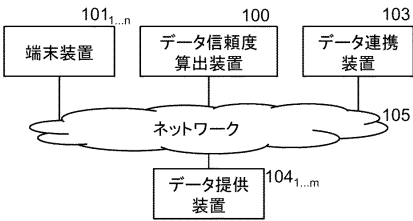
30

40

50

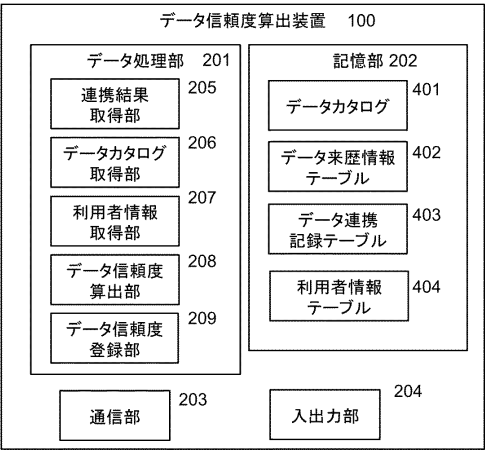
【図面】  
【図 1】

図 1



【図 2】

図 2

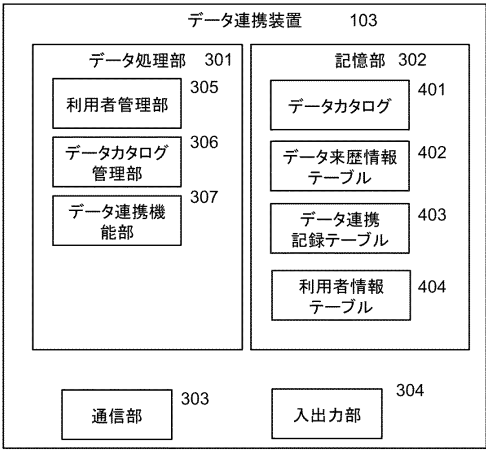


10

20

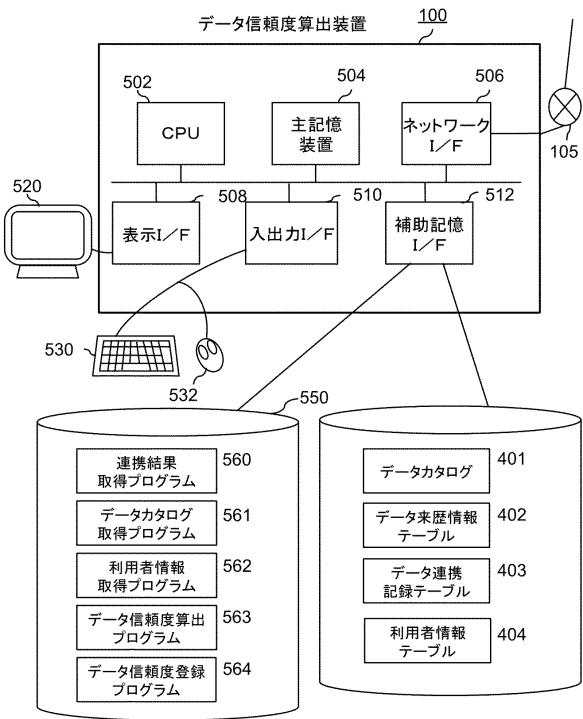
【図 3】

図 3



【図 4】

図 4

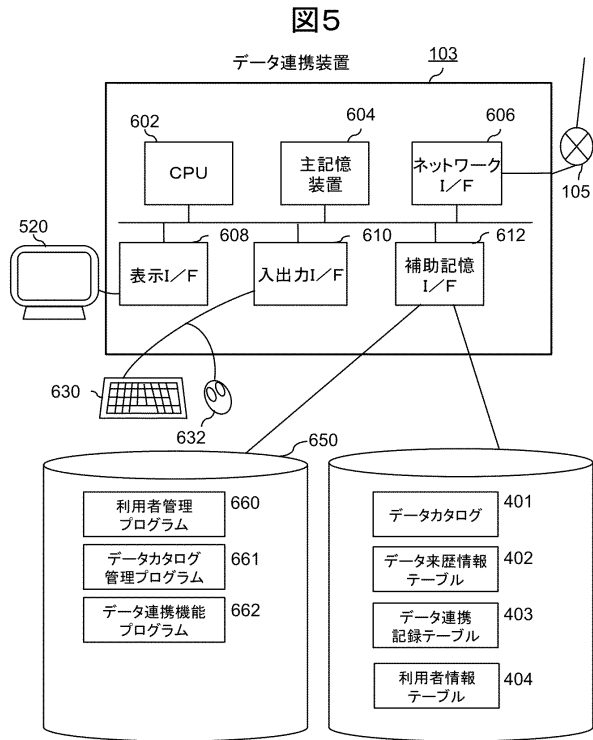


30

40

50

【 図 5 】



【 図 6 】

図6

データカタログ 401

401a DataID	401b Title	401c Trust_Score	401d Sub_Trust_Score	401e Provider
D001	Map	0.427	0.423	P001
D002	Traffic jam	0.167	0.423	P001
D003	Weather	0.190	0.117	P005
D004	Temperature	0.132	0.188,0.190	P003
D005	Kanagawa Weather	0.08	0.271,0.190	P002

10

20

【 図 7 】

図7

データ来歴情報テーブル 402

402a DataID	402b Input_Data_Info	402c Relation
D004	D003	Analysis1
D005	D003	Extracts

【 図 8 】

図8

データ連携記録テーブル 403

403a DataID	403b UserID	403c Relation	403d Date
D001	P001	Provide	2019-9-1
D002	P001	Provide	2019-9-2
D001	U001	Use	2019-9-4
D001	U002	Use	2019-9-4

30

40

50

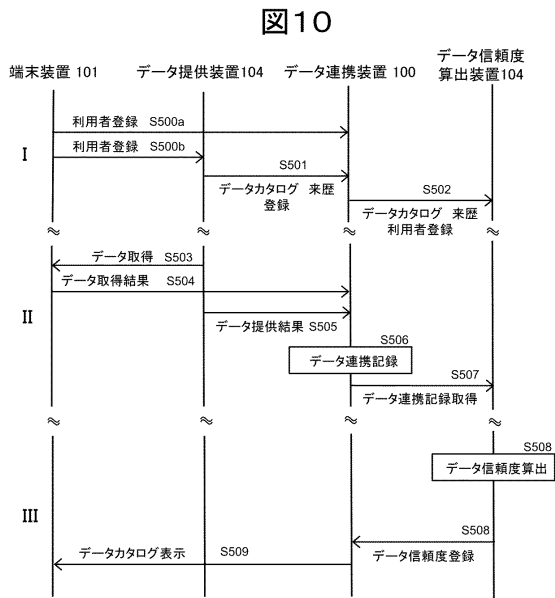
【 図 9 】

図9

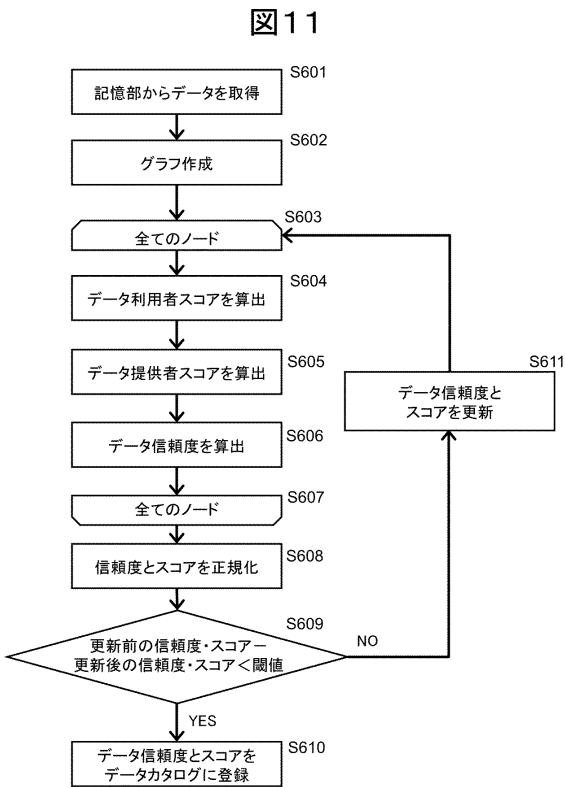
利用者情報テーブル 404

404a	404b	404c
UserID	Name	Organization
U001	Alice	AAA Ltd.
U002	Bob	BBB inc.
U003	Charlie	-
P001	Dave	Ministry of DDD

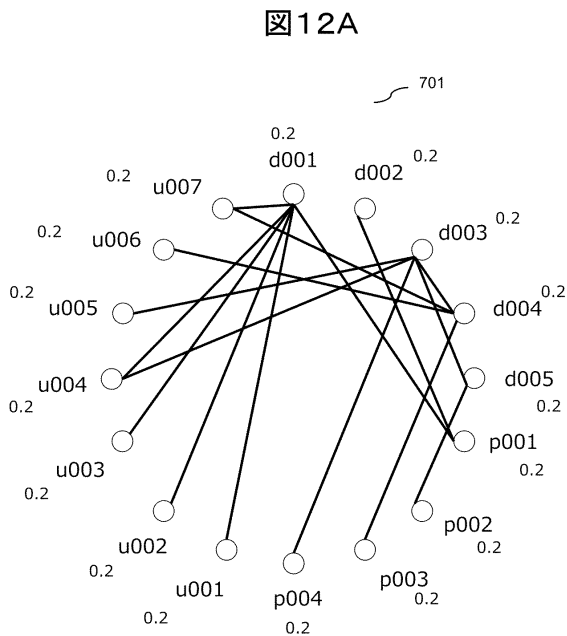
【 図 1 0 】



【 図 1 1 】



【 図 1 2 A 】



10

20

30

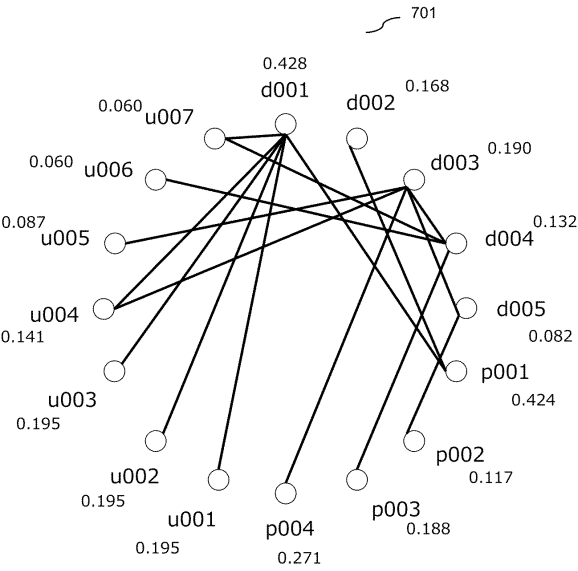
40

50



【 図 1 2 B 】

図12B



10

20

30

40

50

---

フロントページの続き

- (56)参考文献      特開 2 0 0 6 - 2 6 6 7 6 0 ( J P , A )  
                    特開 2 0 0 3 - 3 1 6 6 8 6 ( J P , A )  
                    特開 2 0 0 2 - 0 7 7 1 3 4 ( J P , A )  
                    特開 2 0 1 8 - 0 5 4 5 0 3 ( J P , A )  
                    国際公開第 2 0 1 9 / 2 3 9 4 5 6 ( W O , A 1 )  
                    特開 2 0 0 7 - 1 8 3 8 2 5 ( J P , A )
- (58)調査した分野 (Int.Cl. , D B 名)
- G 0 6 F    1 6 / 0 0 - 1 6 / 9 5 8  
                    G 0 6 Q    1 0 / 0 0 - 9 9 / 0 0