(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7106643号
(P7106643)

(45)発行日　令和4年7月26日(2022.7.26)

(24)登録日　令和4年7月15日(2022.7.15)

(51)国際特許分類
Ｇ０６Ｆ　21/62　(2013.01)

ＦＩ
Ｇ０６Ｆ　21/62　３５４

請求項の数　10　（全23頁）

(21)出願番号　特願2020-531745(P2020-531745)
(86)(22)出願日　平成30年11月29日(2018.11.29)
(65)公表番号　特表2021-507360(P2021-507360A)
(43)公表日　令和3年2月22日(2021.2.22)
(86)国際出願番号　PCT/IB2018/059453
(87)国際公開番号　WO2019/116137
(87)国際公開日　令和1年6月20日(2019.6.20)
審査請求日　令和3年4月23日(2021.4.23)
(31)優先権主張番号　15/843,049
(32)優先日　平成29年12月15日(2017.12.15)
(33)優先権主張国・地域又は機関
米国(US)

(73)特許権者　390009531
インターナショナル・ビジネス・マシーンズ・コーポレーション
ＩＮＴＥＲＮＡＴＩＯＮＡＬ　ＢＵＳＩＮＥＳＳ　ＭＡＣＨＩＮＥＳ　ＣＯＲＰＯＲＡＴＩＯＮ
アメリカ合衆国１０５０４　ニューヨーク州　アーモンク　ニュー　オーチャード　ロード
New Orchard Road, Armonk, New York 10504, United States of America
(74)代理人　100112690
弁理士　太佐　種一

最終頁に続く

(54)【発明の名称】　データを非特定化する方法、データを非特定化するためのシステム、および非データを特定化するためのコンピュータ・プログラム

(57)【特許請求の範囲】

【請求項１】

プロセッサを含むコンピュータ・システムによりデータを非特定化する方法であって、

前記プロセッサを介して、

データセットの実体を識別する1つまたは複数の識別子を決定することと、

前記決定された1つ又は複数の識別子に関連付けられた1つ又は複数のデータ非特定化プロセスを識別することであって、各データ非特定化プロセスが、前記データセット内の保存すべき情報を示す構成の選択肢の１つまたは 複数のセットに関連付けられている、前記識別することと、

前記構成の関連付けられた構成の選択肢のセットに従って、前記データセットに対して前記識別されたデータ非特定化プロセスを実行し、保存される情報が異なるデータセットを生成することと、

前記生成されたデータセットの2つ以上の属性を統合された属性に置き換えることであって、前記統合された属性は、2つ以上の属性のうち、より正確または詳細な情報を含む、前記置き換えることと、

前記生成されたデータセットのプライバシーの脆弱性を評価することと、

前記評価に基づいて、プライバシーの脆弱性が最も少ない生成されたデータセットを生成するデータ非特定化プロセス及び関連付けられた構成の選択肢のセットを選択することと、

前記選択されたデータ非特定化プロセスを、関連付けられた構成の選択肢のセットに従って、前記データセットに対して実行し、結果として非特定化データセットを生成すること

と、
を含む、方法。
【請求項２】
前記１つまたは複数の識別子を決定することが、
１つまたは複数の直接識別子を決定することをさらに含み、前記関連付けられたデータ非特定化プロセスがデータ・マスキング・プロセスを含む、請求項１に記載の方法。
【請求項３】
前記１つまたは複数の識別子を決定することが、
複数の準識別子を決定することをさらに含み、前記関連付けられたデータ非特定化プロセスがデータの一般化またはデータの抑制を含む、請求項１または２に記載の方法。
【請求項４】
前記生成されたデータセットがテーブルの形態であり、前記識別されたデータ非特定化プロセスを実行することが、
生成されたデータセットの２つ以上の列を統合して、前記２つ以上の列より詳細な情報を含む列を生成することをさらに含む、請求項１乃至３のいずれかに記載の方法。
【請求項５】
プライバシーの脆弱性に関して前記生成されたデータセットを評価することが、
生成されたデータセット内の実体のデータと公開されているデータセット内の既知の実体のデータとの間のリンクの存在を決定して、前記生成されたデータセットのプライバシーの脆弱性を示すことをさらに含む、請求項１乃至４のいずれかに記載の方法。
【請求項６】
プライバシーの脆弱性に関して前記生成されたデータセットを評価することが、
対応するデータ非特定化プロセスおよび構成の選択肢の関連付けられたセットによって導入された、生成されたデータセット内の準識別子のセットの存在を決定して、前記生成されたデータセットのプライバシーの脆弱性を示すことをさらに含む、請求項１乃至５のいずれかに記載の方法。
【請求項７】
データ非特定化プロセスごとに一連のテンプレートを生成することをさらに含み、各テンプレートが、前記データ非特定化プロセスの構成の選択肢の関連付けられたセットを指定する、請求項１乃至６のいずれかに記載の方法。
【請求項８】
生成されたデータセットにプライバシーの脆弱性がないことを識別し、前記識別された生成されたデータセットより一般化され情報を含むデータセットを生成する、対応するデータ非特定化プロセスの構成の選択肢の他の関連付けられたセットに関する処理を終了することによって、前記非特定化の処理時間を減らすことをさらに含む、請求項１乃至７のいずれかに記載の方法。
【請求項９】
請求項１乃至７に記載の何れか１項に記載の方法を、コンピュータ・ハードウェアによって実行する、システム。
【請求項１０】
請求項１乃至７に記載の何れか１項に記載の方法を、コンピュータに実行させる、コンピュータ・プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明の実施形態は、データ・アクセスに関連しており、より詳細には、プライバシーおよびデータ有用性を維持しながら非特定化されたデータセットを生成するデータ非特定化プロセスの許容できる構成の検出に基づいてデータを非特定化することに関連している。
【背景技術】
【０００２】

プライバシーを保ちながらデータを公開するプロセスは、直接識別子の発見、直接識別子のマスキング、準識別子（ＱＩＤ：quasi-identifiers）の発見、データ匿名化手法による準識別子の保護、ならびにデータの公開および報告を含む、複数のステップから成る。直接識別子は、実体を直接、一意に識別するために単独で使用されることがある属性であり、一方、準識別子は、実体を一意に識別するために集合的に使用されることがある属性のグループである。上記のプロセス内の異なるステップは、協調して、十分に匿名化されたデータセットが提供されるかどうかを制御する。

【０００３】

データセット内の直接識別子の保護は、データ・マスキング動作を介して実行される。これらの動作は、元のデータ値を新しい架空化されたデータ値に変換し、これらの架空化されたデータ値は、対応する実体を識別するために使用できなくなるが、元のデータ値の特定の情報を保つように特別に作成されてもよく、したがって、データセット内のデータ有用性のレベルを維持することができる。例えば、個人の名前がマスクされるか、または個人の性別情報との一貫性を維持する架空の名前に置き換えられてよく、電子メール（Ｅメール）アドレスがマスクされるか、または元のＥメール・アドレスのドメイン名情報を維持する別のＥメール・アドレスに置き換えられてよく、クレジット・カード番号がマスクされるか、または元のクレジット・カード番号のクレジット・カード発行者情報を反映する別のクレジット・カード番号に置き換えられてよく、電話番号またはファクス番号あるいはその両方がマスクされるか、または元の電話番号またはファクス番号あるいはその両方の国番号または市外局番あるいはその両方を含んでいる別の電話番号またはファクス番号あるいはその両方に置き換えられてよく、郵便番号、市町村、郡、国、および大陸が、元の位置への空間的近接（すなわち、元の値との地理的相関関係）を維持する方法でマスクされてよく、個人に関連する日付がマスクされるか、または元の日付の週数と年、月と年、四半期と年、もしくは年に含まれる別の日付に置き換えられてよく、したがって、複数の医療事例研究などにおいて、特定の種類のその後のデータ解析に非常に役立つ可能性がある極めて重要な情報を維持する。

【０００４】

データセット内の準識別子の保護は、通常、データ一般化動作またはデータ抑制動作を介して実行される。通常、プライバシーを保ちながらデータを公開することにおいて、直接識別子の保護および準識別子の保護は別々に実行される。直接識別子の保護は、有用性を最小限に保って、または有用性を保たずに（例えば、元のデータ値の情報を何も維持しない架空の値に置き換えて）実行され、データの専門家／データ所有者の判断に全体的に基づく。そのような場合、データの専門家／データ所有者は、得られるデータセットが、対象の再特定化、機密情報の開示、会員情報の開示、推論による開示などのプライバシー攻撃に対して十分に保護されるような方法で、データセット内の直接識別子をマスクする方法を決定する必要がある。問題は、直接識別子をマスクするために選択される有用性を保つ選択肢と、データ一般化手法を介して準識別子を保護するために選択される選択肢との間の可能性のある競合に関係している。

【０００５】

直接識別子の新しい値が準識別子の一般化された（新しい）値と一緒に考慮される場合、特定の直接識別子の変換（またはマスキング）において維持される有用性（または情報）は、依然としてプライバシー侵害を許す可能性がある。

【発明の概要】

【発明が解決しようとする課題】

【０００６】

データセットの実体を識別する１つまたは複数の識別子を決定するデータを非特定化するためのシステムを提供する。

【課題を解決するための手段】

【０００７】

本発明の一実施形態によれば、システムがデータを非特定化し、少なくとも１つのプロセ

ッサを備える。このシステムは、データセットの実体を識別する１つまたは複数の識別子を決定する。１つまたは複数のデータ非特定化プロセスが識別され、決定された１つまたは複数の識別子に関連付けられる。各データ非特定化プロセスは、データセット内の保つべき情報を示す構成の選択肢の１つまたは複数のセットに関連付けられる。構成の選択肢の関連付けられたセットに従って、識別されたデータ非特定化プロセスがデータセットに対して実行され、変化する保たれた情報を含むデータセットを生成する。生成されたデータセットが、プライバシーの脆弱性に関して評価され、この評価に基づいて、データ非特定化プロセスおよび構成の選択肢の関連付けられたセットが選択される。構成の選択肢の関連付けられたセットに従って、選択されたデータ非特定化プロセスがデータセットに対して実行され、結果として得られる非特定化されたデータセットを生成する。本発明の実施形態は、前述したのと実質的に同じやり方でデータを非特定化するための方法およびコンピュータ・プログラム製品をさらに含む。

【０００８】

本発明の実施形態は、データを非特定化するためのデータ非特定化プロセスを選択するために試行錯誤手法を採用するのではなく、データ非特定化プロセスの実行可能な構成または最適な構成あるいはその両方を識別することによって、処理時間を減らす。これらの試行錯誤による選択は、通常、ユーザの知識に基づき、準最適なデータ非特定化および多数のデータ非特定化の試行につながることがあり、それによって、処理リソースおよびその他のリソースを浪費する。

【０００９】

本発明の実施形態は、テーブルの形態で評価用のデータセットをさらに生成し、生成されたデータセットの２つ以上の列を統合して、それら２つ以上の列より詳細な情報を含む列を生成することができる。これによって、より詳細な情報を含むデータセットを評価して、プライバシーの脆弱性がないことを保証できるようにする。より詳細な情報を含む生成されたデータセットにプライバシーの脆弱性がない場合、さらに一般化された情報を含む、対応するデータ非特定化プロセスおよび構成の選択肢から生成された他のデータセット（例えば、元の統合されていない列のうちの１つまたは複数を含むデータセット）にもプライバシーの脆弱性がなくなる。これによって、より詳細な情報および一般化された情報を含むデータセットの複数の評価の代わりに、単一の評価を利用して、処理時間も削減する。

【００１０】

本発明の実施形態は、生成されたデータセット内の実体のデータと公開されているデータセット内の既知の実体のデータとの間のリンクの存在を決定して、生成されたデータセットのプライバシーの脆弱性を示すことによって、プライバシーの脆弱性に関して生成されたデータセットを評価することができる。この評価は、公開されているデータセット内の既知の実体に対して、生成されたデータセットからの非特定化されたデータを利用して、非特定化されたデータ内の実体の同一性が三角測量攻撃（triangulation attacks）によって決定され得るかどうかを判定し、それによって、関連付けられた構成の選択肢を有する推奨されたデータ非特定化プロセスがプライバシーを維持するという高い信頼性を実現する。

【００１１】

本発明の実施形態は、対応するデータ非特定化プロセスおよび構成の選択肢の関連付けられたセットによって導入された、生成されたデータセット内の準識別子のセットの存在を決定して、生成されたデータセットのプライバシーの脆弱性を示すことによって、プライバシーの脆弱性に関して生成されたデータセットを評価することができる。この評価は、一意性の基準に基づき、準識別子が、データ非特定化プロセスおよび関連付けられた構成の選択肢によって導入されないということを保証し、それによって、関連付けられた構成の選択肢を有する推奨されたデータ非特定化プロセスがプライバシーを維持するという高い信頼性を実現する。生成されたデータセットが一意性も外れ値も含んでいない場合、そのデータセットは、三角測量攻撃によってどの他の（内部または外部の）データセットに

もリンクされ得ず、したがってプライバシーを維持する。
【００１２】
本発明の実施形態は、生成されたデータセットにプライバシーの脆弱性がないことを識別し、識別された生成されたデータセットより一般化され情報を含むデータセットを生成する、対応するデータ非特定化プロセスの構成の選択肢の他の関連付けられたセットに関する処理を終了することによって、元のデータセットの非特定化の処理時間を減らすことができる。これによって、プロセッサの性能を大幅に改善し、処理時間を短縮することにおいて最適なデータ非特定化を実現する。
【００１３】
通常、さまざまな図内の類似する参照番号は、類似するコンポーネントを指定するために利用される。
【図面の簡単な説明】
【００１４】
【図１】本発明の実施形態の例示的なコンピューティング環境の概略図である。
【図２】本発明の実施形態による、データのプライバシーを維持してデータセットを生成するための、データ非特定化プロセスの許容できる構成の選択肢を検出する方法を示す手順のフローチャートである。
【図３】本発明の実施形態による、データ非特定化プロセスの構成の選択肢に従ってデータセットを生成する方法の手順のフローチャートである。
【図４】本発明の実施形態による、公開されているデータに基づいてデータ非特定化プロセスの構成の選択肢を評価する方法の手順のフローチャートである。
【図５】本発明の実施形態による、非特定化されたデータ内の準識別子の導入に基づいてデータ非特定化プロセスの構成の選択肢を評価する方法の手順のフローチャートである。
【図６】性別情報を保ちながら名前属性を非特定化するように構成されたデータ非特定化プロセスによって生成された例示的なデータセットを示す図である。
【図７】空間的近接を保ちながら住所属性を非特定化するように構成されたデータ非特定化プロセスによって生成された例示的なデータセットを示す図である。
【図８】本発明の実施形態による、処理時間を減らすようにデータ非特定化プロセスの処理を制御するために利用される例示的なツリー構造の概略図である。
【発明を実施するための形態】
【００１５】
本発明の実施形態は、データ非特定化プロセスまたは手法の構成の選択肢の各使用可能なセットのプライバシーのリスクを評価し、データ内のプライバシーの脆弱性を阻止する構成の選択肢（または設定）のみを使用できるようにする。本発明の実施形態は、データセットを解析し、データ匿名化を実行するためのデータ非特定化プロセスまたは手法の許容される構成の選択肢（または設定）を発見して報告する。構成の選択肢または設定は、通常、非特定化されるデータおよび非特定化された値によって保たれるデータ内の対応する情報を示す。例として、本発明の実施形態は、データセットの直接識別子のデータ・マスキング・プロセスまたは手法のための構成の選択肢を検出してよい。しかし、任意のデータ非特定化もしくは匿名化プロセスまたは手法が、下で説明される方法と実質的に同じ方法で、任意の種類の識別子に関して評価されてよい。
【００１６】
既存の方法では、データを非特定化するためのデータ非特定化プロセスを選択するために、試行錯誤手法が通常は採用される。これらの選択は、通常、ユーザの知識に基づき、準最適なデータ非特定化および多数のデータ非特定化の試行につながることがあり、それによって、処理リソースおよびその他のリソースを浪費する。本発明の実施形態は、有用性を最大限に保つ方法においてデータを迅速に非特定化するために、データ非特定化プロセスの許容できる構成または最適な構成あるいはその両方を識別することによって、処理時間を短縮する。
【００１７】

本発明の一実施形態によれば、データセットの実体を識別する１つまたは複数の識別子（属性）が決定される。１つまたは複数のデータ非特定化プロセスが識別され、決定された１つまたは複数の識別子に関連付けられる。各データ非特定化プロセスは、保つべき情報を示す（有用性を保つ）構成の選択肢の１つまたは複数のセットに関連付けられる。データセット内の識別子ごとに、有用性を保つ構成を有するデータ非特定化プロセスが選択される。識別子のデータ非特定化プロセスのうちで、識別子を完全に抑制する特別な場合について考える。構成の選択肢の関連付けられたセットに従って、選択されたデータ非特定化プロセスがデータセットに対して実行され、変化するデータ有用性が保たれたデータセットを生成する。その後、少なくとも１つの識別子に関して、有用性を保つ構成と共に、異なるデータ非特定化プロセスが選択され、構成の選択肢の関連付けられたセットに従って、新たに選択されたデータ非特定化プロセスがデータセットに対して実行され、変化するデータ有用性が保たれた新しいデータセットを生成する。データセットの実体を識別する決定された１つまたは複数の識別子に関して、異なるデータ非特定化プロセスのすべての可能性のある組み合わせ、およびそれらに関連付けられた構成の選択肢がデータセットに対して実行されるまで、同じ動作が繰り返され、変化するデータ有用性が保たれたデータセットを生成する。各生成されたデータセットが、プライバシーの脆弱性に関して評価され、この評価に基づいて、１つまたは複数のデータ非特定化プロセスおよび構成の選択肢の関連付けられたセットが選択される。選択されたデータ非特定化プロセスのうちで、最低の再特定化のリスクおよび最高のデータ有用性を実現するデータ非特定化プロセスが、構成の選択肢の関連付けられたセットに従って、データセットに対して実行され、結果として得られる非特定化されたデータセットを生成する。

【００１８】

本発明の実施形態は、テーブルの形態で評価用のデータセットをさらに生成し、生成されたデータセットの２つ以上の列を統合して、それら２つ以上の列より詳細な情報を含む列を生成することができる。これによって、より詳細な情報を含むデータセットを評価して、プライバシーの脆弱性がないことを保証できるようにする。より詳細な情報を含む生成されたデータセットにプライバシーの脆弱性がない場合、さらに一般化された情報を含む、対応するデータ非特定化プロセスおよび構成の選択肢から生成された他のデータセット（例えば、元の統合されていない列のうちの１つまたは複数を含むデータセット）にもプライバシーの脆弱性がなくなる。これによって、より詳細な情報および一般化された情報を含むデータセットの複数の評価の代わりに、単一の評価を利用して、処理時間も削減する。

【００１９】

さらに、本発明の実施形態は、生成されたデータセット内の実体のデータと公開されているデータセット内の既知の実体のデータとの間のリンクの存在を決定して、生成されたデータセットのプライバシーの脆弱性を示すことによって、プライバシーの脆弱性に関して生成されたデータセットを評価することができる。この評価は、公開されているデータセット内の既知の実体に対して、生成されたデータセットからの非特定化されたデータを利用して、非特定化されたデータ内の実体の同一性が決定され得るかどうかを判定し、それによって、関連付けられた構成の選択肢を有する推奨されたデータ非特定化プロセスがプライバシーを維持するという高い信頼性を実現する。

【００２０】

本発明の実施形態は、対応するデータ非特定化プロセスおよび構成の選択肢の関連付けられたセットによって導入された、生成されたデータセット内の準識別子のセットの存在を決定して、生成されたデータセットのプライバシーの脆弱性を示すことによって、プライバシーの脆弱性に関して生成されたデータセットをさらに評価することができる。この評価は、準識別子が、データ非特定化プロセスおよび関連付けられた構成の選択肢によって導入されないということを保証し、それによって、関連付けられた構成の選択肢を有する推奨されたデータ非特定化プロセスがプライバシーを維持するという高い信頼性を実現する。

【００２１】
加えて、本発明の実施形態は、生成されたデータセットにプライバシーの脆弱性がないことを識別し、識別された生成されたデータセットより一般化され情報を含むデータセットを生成する、対応するデータ非特定化プロセスの構成の選択肢の他の関連付けられたセットに関する処理を終了することによって、非特定化の処理時間を減らすことができる。これによって、プロセッサの性能を大幅に改善し、処理時間を短縮することにおいて最適なデータ非特定化を実現する。
【００２２】
本発明の実施形態で使用するための例示的な環境が、図１に示されている。具体的には、この環境は、１つまたは複数のサーバ・システム１１０および１つまたは複数のクライアントもしくはエンドユーザ・システム１１４を含んでいる。サーバ・システム１１０およびクライアント・システム１１４は、互いにリモートの位置にあり、ネットワーク１１２を経由して通信してよい。このネットワークは、任意の数の任意の適切な通信媒体（例えば、広域ネットワーク（ＷＡＮ：wide area network）、ローカル・エリア・ネットワーク（ＬＡＮ：local area network）、インターネット、イントラネットなど）によって実装されてよい。代替として、サーバ・システム１１０およびクライアント・システム１１４は、互いにローカルな位置にあり、任意の適切なローカル通信媒体（例えば、ローカル・エリア・ネットワーク（ＬＡＮ）、ハードワイヤ、無線リンク、イントラネットなど）を介して通信してよい。
【００２３】
クライアント・システム１１４は、ユーザがサーバ・システム１１０と情報をやりとりして、データ非特定化などの望ましい動作を実行できるようにする。サーバ・システムは、さまざまなデータ非特定化プロセスもしくは手法の許容できる構成または設定を検出し、データのプライバシーを維持する結果として得られるデータセットを生成するために、評価モジュール１１６を含む。データベース・システム１１８は、解析用のさまざまな情報（例えば、元のデータセットおよび暫定的なデータセット、構成または設定、データ非特定化プロセスの選択肢など）を格納してよい。データベース・システムは、任意の従来もしくはその他のデータベースまたはストレージ・ユニットによって実装されてよく、サーバ・システム１１０およびクライアント・システム１１４からローカルまたはリモートの位置にあってよく、任意の適切な通信媒体（例えば、ローカル・エリア・ネットワーク（ＬＡＮ）、広域ネットワーク（ＷＡＮ）、インターネット、ハードワイヤ、無線リンク、イントラネットなど）を介して通信してよい。クライアント・システムは、グラフィカル・ユーザ・インターフェイス（例えば、ＧＵＩなど）またはその他のインターフェイス（例えば、コマンド・ライン・プロンプト、メニュー画面など）を提示して、解析に関するユーザからの情報を求めてよく、解析結果を含んでいる報告（例えば、推奨されるデータ非特定化プロセス、非特定化されたデータセット、データセットを非特定化するために使用された選択肢など）を提供してよい。
【００２４】
サーバ・システム１１０およびクライアント・システム１１４は、ディスプレイまたはモニタ、基盤、任意選択的な入力デバイス（例えば、キーボード、マウス、またはその他の入力デバイス）、ならびに任意の市販のソフトウェアおよびカスタム・ソフトウェア（例えば、サーバ／通信ソフトウェア、評価モジュール、ブラウザ／インターフェイス・ソフトウェア、データ非特定化プロセスなど）を備えているのが好ましい、任意の従来またはその他のコンピュータ・システムによって実装されてよい。基盤は、少なくとも１つのハードウェア・プロセッサ１１５（例えば、マイクロプロセッサ、コントローラ、中央処理装置（ＣＰＵ：central processing unit）など）、１つまたは複数のメモリ１３５、または内部もしくは外部ネットワーク・インターフェイスもしくは通信デバイス１２５（例えば、モデム、ネットワーク・カードなど）、あるいはその組み合わせを含んでいるのが好ましい。
【００２５】

代替として、１つまたは複数のクライアント・システム１１４は、スタンドアロン・ユニットとして動作しているときに、さまざまなデータ非特定化プロセスまたは手法の許容できる構成または設定を検出してよい。スタンドアロン・モードの動作では、クライアント・システムが、データ（例えば、データセット、構成または設定、データ非特定化プロセスなど）を格納するか、またはデータにアクセスすることができ、検出を実行するための評価モジュール１１６を含む。グラフィカル・ユーザ・インターフェイス（例えば、ＧＵＩなど）またはその他のインターフェイス（例えば、コマンド・ライン・プロンプト、メニュー画面など）は、解析に関する対応するユーザからの情報を求め、解析結果を含んでいる報告を提供してよい。

【００２６】

評価モジュール１１６は、以下で説明されている本発明の実施形態のさまざまな機能を実行するために、１つまたは複数のモジュールまたはユニットを含んでよい。任意の数のソフトウェア・モジュールもしくはユニットまたはハードウェア・モジュールもしくはユニットあるいはその両方の任意の組み合わせによって、さまざまなモジュール（例えば、評価モジュールなど）が実装されてよく、プロセッサ１１５によって実行するために、サーバ・システムまたはクライアント・システムあるいはその両方のメモリ１３５内に存在してよい。

【００２７】

本発明の実施形態に従って、データ非特定化プロセスもしくは手法の許容できる構成の選択肢または設定を（例えば、評価モジュール１１６およびサーバ・システム１１０またはクライアント・システム１１４あるいはその両方を介して）検出し、データのプライバシーを維持してデータセットを生成する方法が、図２に示されている。最初に、各データ非特定化プロセスが、特定のデータの種類に関連付けられ、さまざまな構成の選択肢または設定に従って動作する。構成の選択肢または設定は、通常、非特定化されるデータおよび非特定化された値によって保たれるデータ内の対応する情報を示す。例えば、性別情報を保ちながら非特定化される名前を指定する構成の選択肢は、元の性別情報と一致する架空化された名前に置き換えられた名前を含むデータセットを生成する（例えば、元の名前などの性別情報との一貫性を保つまたは維持するように、女性の名前が架空化された女性の名前に置き換えられる）。加えて、この構成の選択肢は、特定のデータが結果として得られるデータセットから削除されることを指定してよい。

【００２８】

データ非特定化プロセスに関連付けられた構成の選択肢のセットごとに、テンプレートが生成される。構成の選択肢の各セットは、データ非特定化プロセスの１つまたは複数の構成の選択肢を含んでよい。したがって、（サーバ・システムまたはクライアント・システムあるいはその両方で使用可能な）各データ非特定化プロセスが一連のテンプレートに関連付けられ、各テンプレートが、そのデータ非特定化プロセスの構成の選択肢の可能性のあるセットのうちの１つ（例えば、削除されるデータ、他のデータを保ちながら非特定化されるデータ、空間的近接などの特定の特性を保ちながら非特定化されるデータなど）に対応する。基本的に、データ非特定化プロセスごとの一連のテンプレートは、関連付けられた属性または識別子に関するそのデータ非特定化プロセスのすべての可能性のある構成を対象にする。対応する構成の選択肢に従って、関連付けられたデータ非特定化プロセスによって元の属性が処理されるときに、テンプレートが、データセット内で保持されている情報を捕捉する。

【００２９】

例えば、名前属性、電話番号属性、および住所属性に関するデータ非特定化プロセスのテンプレートは、名前テンプレート（例えば、名前属性が削除されるテンプレート（名前、削除）、名前属性が性別情報との一貫性を保つまたは維持する値に置き換えられるテンプレート（名前、性別））、電話テンプレート（例えば、電話番号属性が削除されるテンプレート（電話、削除）、電話番号属性が国、ならびに国番号および市外局番との一貫性をそれぞれ保つまたは維持する値に置き換えられるテンプレート（電話、国）、テンプレー

ト（電話、国および地域））、住所テンプレート（例えば、住所属性が削除されるテンプレート（住所、削除）、住所属性が国、市町村、および規定の距離の範囲内の地域との一貫性をそれぞれ保つまたは維持する値に置き換えられるテンプレート（住所、国）、テンプレート（住所、国および市町村）、テンプレート（住所、最小境界矩形（ＭＢＲ：minimum bounding rectangle））を含んでよい。しかし、テンプレートは、任意の属性（例えば、住所、電話番号、車両識別番号（ＶＩＮ：vehicle identification number）、社会保障番号（ＳＳＮ：social security number）、国、ユニフォーム・リソース・ロケータ（ＵＲＬ：uniform resource locator）、名前、ＩＰアドレス、電子メール（Ｅメール）アドレス、クレジット・カード番号、国際銀行番号（ＩＢＡＮ：international bank account number）、日付、市町村、医療ＩＣＤコード（medical ICD code）、職業、病院、緯度／経度、郵便番号など）を削除するまたは保つための任意の望ましい選択肢に関連してよい。データのプライバシーおよびデータ有用性の保存に関して、テンプレートは、非特定化の後にデータセット内で維持される真実の情報を捕捉する。テンプレート（属性Ａ、選択肢Ｂ）の場合、このテンプレートは、データセット内の属性Ａを選択肢Ｂで提供された（有用性を保つ）情報に置き換えることを表す。例えば、テンプレート（名前、性別）は、データセット内の名前属性を、データ内の個人に関する正確な性別情報を捕捉する性別属性に置き換えることとして変換され得る。同様に、テンプレート（電話、国および地域）は、データセット内の電話属性を、データセット内で表された個人に関する正確な国情報を維持する属性および正確な地域情報を維持する属性に置き換えることとして変換され得る。テンプレートの使用は、データ内で何が保持されているかに関する情報を提供し、この情報は、その後、結果として得られるデータセット内のプライバシーのリスクおよびデータ有用性を計算するために使用され得る。

【００３０】

加えて、テンプレートは、削除もしくは非特定化する１つもしくは複数の属性、または保つべき１つもしくは複数の属性、あるいはその両方を示してよい。例えば、一連のテンプレートは、構成の選択肢に従って削除または非特定化する属性をそれぞれ指定する初期テンプレートを含んでよい。追加のテンプレートは、初期テンプレートまたは属性の構成の選択肢を指定し、第２の属性に関する構成の選択肢をさらに含んでよい（例えば、２つの属性の非特定化を提供する）。したがって、データ非特定化プロセスのテンプレートは、データセットの対応する属性に関する、データ非特定化プロセスによって提供される非特定化のさまざまな組み合わせのすべてまたは任意の部分を対象にしてよい。

【００３１】

例として、本発明の実施形態は、データセットの直接識別子のデータ・マスキング・プロセスまたは手法の形態で、データ非特定化プロセスの構成の選択肢を検出することに関して説明される。しかし、任意のデータ非特定化もしくは匿名化プロセスまたは手法が、下で説明される方法と実質的に同じ方法で、任意の種類の識別子に関して評価されてよい。

【００３２】

具体的には、データセット２５０が受信され、ステップ２０５で、データ・マスキング用の直接識別子を検出するために解析される。直接識別子は、実体を直接識別するために使用されることがある属性（例えば、名前、社会保障番号、住所、電話番号など）である。データセットは、各行が実体を表し、各列がその実態の属性（例えば、名前、住所、性別など）を表しているテーブルの形態であることが好ましい。しかし、データセットは任意の望ましい形式であってよい。直接識別子は、任意の従来またはその他の手法を使用して検出されてよい。例えば、実体に関する属性の一意性が、データセット２５０内の直接識別子を検出するために使用されてよい。代替として、規則的な表現またはパターンが、直接識別子であることが知られているデータセット内の特定の種類のデータ（例えば、社会保障番号、住所、日付など）を識別するために使用されてよい。代替として、ルックアップ・テーブルが、（例えば、有権者登録リストを介して）名前などの特定の種類の直接識別子を識別するために使用されてよい。加えて、データセットの直接識別子は、ユーザによって手動で事前に決定されてよい。

【００３３】
ステップ２１０で、検出された直接識別子に対応するデータ・マスキング・プロセスが識別される。データ・マスキング・プロセスは、通常、特定の種類のデータまたは属性に適合し、各検出された直接識別子が、評価のために、対応する適合するデータ・マスキング・プロセスの各々に関連付けられる。
【００３４】
ステップ２１５で、データ・マスキング・プロセスが、データ・マスキング・プロセスの構成の選択肢のさまざまなセットを指定する（前述した）テンプレートに従って、対応する直接識別子に適用される。これによって、直接識別子に関連付けられたデータ・マスキング・プロセスごとに、構成の選択肢の各セットのデータセットを生成する。生成されるデータセットは、行および列（または属性）のテーブルの形態であることが好ましいが、任意の望ましい形式であってよい。例えば、図６は、個人を表す各行と、個人ごとの名前、住所、生年月、郵便番号、および結婚歴の列または属性とを含むテーブルの形態で、初期データセット６００を示している。データ・マスキング・プロセスは、性別属性との一貫性を保つまたは維持する架空化された名前を使用して、名前属性をマスクできるようにしてよい。この場合、データ・マスキング・プロセスのテンプレートは、構成の選択肢の対応するセット（例えば、テンプレート（名前、性別））を指定してよい。
【００３５】
この構成の選択肢のセットに従ってデータ・マスキング・プロセスが適用されるときに、個人の名前が性別属性との一貫性を保つまたは維持する架空化された名前を使用してマスクされて、データセット６２０が生成される。これによって、元のデータセット６００から計算された正確な性別情報を含んでいる新しい性別属性が現れているデータセット６２０が、効果的に得られる。この場合、データセット６００内の男性の名前が、性別情報を維持するように、データセット６２０内の異なる男性の名前に置き換えられている。同様に、データセット６００内の女性の名前が、性別情報を保つように、データセット６２０内の異なる女性の名前に置き換えられている。これによって、（架空化された名前が、個人の性別のみを識別し、データ内のプライバシーのリスクを高めるどの他の目的にも使用され得ないため）プライバシーの脆弱性を評価することに関して効果的に、名前属性または名前列を性別列に置き換える。
【００３６】
さらに別の例として、図７は、個人を表す各行と、個人ごとの名前、住所、生年月、郵便番号、および結婚歴の列または属性とを含むテーブルの形態で、初期データセット７００を示している。データ・マスキング・プロセスは、２マイルの最小境界矩形（ＭＢＲ）の範囲内の別の住所を使用して住所属性をマスクできるようにしてよい。この場合、データ・マスキング・プロセスのテンプレートは、構成の選択肢の対応するセット（例えば、テンプレート（住所、最小境界矩形（ＭＢＲ））を指定してよい。
【００３７】
この構成の選択肢のセットに従ってデータ・マスキング・プロセスが適用されるときに、個人の住所が２マイルの最小境界矩形（ＭＢＲ）の範囲内にある異なる住所に変更またはマスクされて、データセット７２０が生成される。しかし、この構成の選択肢のセットの場合、郵便番号と組み合わせた新しい住所が準識別子を形成し、プライバシーの脆弱性を作り出すことがある。したがって、生成されたデータセット７２０において、住所属性と郵便番号属性を組み合わせて、個人の位置（例えば、自宅の住所）に関するできるだけ多くの特定性を取得する必要がある。その後、この情報を使用して、データを公開することのプライバシーのリスクを評価する。
【００３８】
再び図２を参照すると、ステップ２２０で、許容できるデータ・マスキング・プロセスおよび構成の選択肢の対応するセットを識別して、データのプライバシーを維持する結果として得られるデータセットを生成するために、テンプレートから生成されたデータセットが評価される。この評価は、公開されているデータセットまたは外部のデータセット（例

えば、有権者登録リスト、職業別電話帳、国勢調査データなど）とのつながりに関して、生成されたデータセットを解析する。つながりが存在する場合（例えば、外部のデータセットを使用した三角測量攻撃が成功する場合）、それは、生成された（またはマスクされた）データセットの個人の身元が決定されることがあるということを示し、それによって、データセットの生成に使用されるデータ・マスキング・プロセスおよび構成の選択肢の対応するセットに関するプライバシーの脆弱性を識別する。加えて、生成されたデータセットが解析され、データ・マスキング・プロセスおよび構成の選択肢の対応するセットに基づいて、生成されたデータセットに導入された準識別子の存在を決定してよい。準識別子の存在は、データセットを生成するために使用されるデータ・マスキング・プロセスおよび構成の選択肢の対応するセットに関するプライバシーの脆弱性を示す。

【００３９】

結果として得られるデータ・マスキング・プロセスおよび構成の選択肢の対応するセットは、識別された許容できるデータ・マスキング・プロセス（および構成の選択肢の対応するセット）の中から選択されてよい。結果として得られるデータ・マスキング・プロセスは、ユーザによって手動で選択されてよい。この場合、許容できるデータ・マスキング・プロセスおよび構成の選択肢の関連付けられたセットが、クライアント・システム１１４上のユーザに、選択のために提示されてよい。許容できるデータ・マスキング・プロセスの推奨が提供されてもよい。この推奨は、さまざまな指標（例えば、プライバシーのレベル、処理時間、データの保存など）に基づいてよい。

【００４０】

代替として、結果として得られるデータ・マスキング・プロセスは、自動的に決定されてよい。結果として得られるデータ・マスキング・プロセスを決定するために、さまざまな指標が利用されてよい。例えば、公開されているデータセットとのつながりまたは最少の数の準識別子の導入あるいはその両方に基づいて、最高のデータのプライバシーを提供するデータ・マスキング・プロセスが選択されてよい。代替として、データセットの非特定化のための処理時間を減らすために、最少のリソース使用量または処理時間あるいはその両方に基づいて、データ・マスキング・プロセスが選択されてよい。

【００４１】

加えて、結果として得られるデータ・マスキング・プロセスは、機械学習に基づいて推奨されるか、または自動的に選択されてよい。この場合、ユーザによって選択されたデータ・マスキング・プロセスおよび構成の選択肢の関連付けられたセットが格納されてよく、または指標が追跡されてよく、あるいはその両方が行われてよい。この情報が処理され、選択または推奨あるいはその両方に関するユーザの嗜好を学習してよい。学習を実行するためのさまざまなモデル（例えば、ニューラル・ネットワーク、数学／統計モデル、分類器など）が採用されてよい。例えば、マスキング・プロセスが最初に推奨されるか、または選択されるか、あるいはその両方が行われてよい。しかし、何らかの理由で、あるユーザは、別の許容できるデータ・マスキング・プロセスを繰り返し好んだ。ユーザのこれらの側面および嗜好が、学習されてよく（例えば、ユーザが、より高いプライバシーのレベルよりも、より短い処理時間を好むことがあるなど）、データ・マスキング・プロセスを選択するか、または推奨するか、あるいはその両方を実行するために、採用されてよい。

【００４２】

ステップ２２５で、データのプライバシーを維持しながらデータセットを非特定化するために、構成の選択肢の対応するセットに従って、結果として得られるデータ・マスキング・プロセスがデータセット２５０に適用される（または、データセット２５０に対して実行される）。

【００４３】

本発明の実施形態に従って、評価用のデータセットを生成するためにデータ・マスキング・プロセスのテンプレートを適用する方法（例えば、図２のステップ２１５に対応する）が、図３に示されている。最初に一連のデータ・マスキング・プロセスおよび構成の選択肢の対応するセットが使用され、可能性のあるプライバシーのリスクの導入に関してテス

トされるデータセットを生成する。具体的には、ステップ３０５で、検出された直接識別子に関連付けられたデータ・マスキング・プロセスごとに、構成の選択肢の異なるセットが決定される。ステップ３１０で、各データ・マスキング・プロセスの構成の選択肢の決定されたセットごとに、データセットが生成される。これは、構成の選択肢のセットを指定するテンプレートをデータ・マスキング・プロセスに適用して、データセットを生成することによって、実現される。言い換えると、テンプレートの構成の選択肢のセットに従ってデータ・マスキング・プロセスが実行され、関連付けられた直接識別子を削除またはマスクする。生成されるデータセットは、行および列（または属性）を含むテーブルの形態であることが好ましいが、任意の望ましい形式であってよい。

【００４４】

ステップ３１５で、同じ種類または適合する種類である生成されたデータセット内の属性または列が統合され、生成されたデータセット内のより正確または詳細な情報を含む列を提供してよい。例えば、統合された列は、統合されている初期の列内の地域または位置の共通集合であってよい。例えば、郵便番号および住所の最小境界矩形（ＭＢＲ）をそれぞれ含んでいる列が、位置に関してより正確な情報を含んでいる単一の列に置き換えられてよい。この場合、ＭＢＲが郵便番号より広い地域をカバーしているときに、位置に関してより詳細な情報を提供する（例えば、郵便番号がＭＢＲより狭い地域をカバーする）ように、郵便番号列が生成されたデータセット内に残ってよい。これによって、プライバシーの脆弱性に関してテストされる、より詳細な情報を含む生成されたデータセット（または、プライバシーの脆弱性の影響をより受けやすい状況）を提供する。より詳細な情報がプライバシーの問題を引き起こさない場合、どんな一般化された情報またはより広い情報も、同様にプライバシーの問題を引き起こさない。

【００４５】

データ・マスキング・プロセスおよび構成の選択肢の関連付けられたセットの各々の生成されたデータセットが、プライバシーの脆弱性に関して評価される。

【００４６】

公開されているデータに基づいて生成されたデータセットのプライバシーの脆弱性を検出する方法（例えば、図２のステップ２２０に対応する）が、図４に示されている。最初に、データ・マスキング・プロセスおよび構成の選択肢のセットを指定する対応するテンプレートから生成された各データセットが、プライバシーの脆弱性に関して評価される。これは、生成されたデータセット内のデータを外部のデータまたは公開されているデータに結び付けることによって実現される。具体的には、ステップ４０５で、各生成されたデータセット内のデータが、外部のデータまたは公開されているデータ（例えば、有権者登録リスト、職業別電話帳、国勢調査データなど）との可能性のあるつながりに関してテストされる。言い換えると、生成されたデータセット内の実体のデータは、公開されているデータ内の対応する既知の実体のデータへのリンクを決定するために利用される。例えば、生成されたデータセット内の実体の１つまたは複数の属性値が、公開されているデータ内の対応する属性値を見つけるために使用されてよい。

【００４７】

リンクが存在する（例えば、十分な数またはパターンの属性が一致する）場合、それは、生成されたデータセットの実体のデータが、公開されているデータ内の既知の実体に対応しており、それによって、生成されたデータセットからの実体の識別を可能にするということを示している。生成されたデータセットの実体と公開されているデータとの間で、ある数のリンクが維持されていることがあり、フロー４１０で、しきい値と比較され、生成されたデータセット（ならびに、生成されたデータセットを生成するために使用されるデータ・マスキング・プロセスおよび構成の選択肢のセット）に関するプライバシーの脆弱性の存在を決定することがある。このしきい値は、任意の望ましい値に設定されてよく、プライバシーの脆弱性を示すための任意の望ましい方法（例えば、より大きい、より小さい、以上、以下など）で、リンクの数がこのしきい値と比較されてよい。例えば、このしきい値が０に設定されてよく、生成されたデータセットの実体と、公開されているデータ

の既知の実体との間の１つまたは複数のリンクの存在に応答して、生成されたデータセットがプライバシーの脆弱性を有していると見なされてよい。プライバシーの脆弱性を有する生成されたデータセットを生成するために使用されるデータ・マスキング・プロセスおよび構成の選択肢の関連付けられたセットが、推奨または選択あるいはその両方を決定するためにマーク付けされる。

【００４８】

生成されたデータセットの各々が外部のデータまたは公開されているデータに対してテストされた後に、プライバシーの脆弱性を有する生成されたデータセットを生成するために使用されるデータ・マスキング・プロセスおよび構成の選択肢の関連付けられたセットがマーク付けされ、それ以上の検討から除外される。ステップ４１５で、残りのデータ・マスキング・プロセスおよび構成の選択肢の関連付けられたセットが解析され、データ・マスキング・プロセスの推奨されるセットおよび構成の選択肢の関連付けられたセットを決定し、脆弱でないデータセットを提供する。保存がより少ない構成の選択肢の関連付けられたセットを有するデータ・マスキング・プロセスを除去することによって、推奨されるセットが減らされてよい。加えて、データ・マスキング・プロセスおよび構成の選択肢の関連付けられたセットが、プライバシーの脆弱性を有さないデータセットを提供しない場合、最も少ないプライバシーの脆弱性（または、例えばリンクの数）を有するデータ・マスキング・プロセスおよび構成の選択肢の関連付けられたセットが推奨されてよい。前述したように、推奨されるデータ・マスキング・プロセスが、選択のためにユーザに提示されてよく、またはデータ・マスキング・プロセスが自動的に選択されてよい。

【００４９】

加えて、図５に示されているように、生成されたデータセットの解析（例えば、図２のステップ２２０に対応する）に基づいて、生成されたデータセットのプライバシーの脆弱性が決定されてよい。最初に、ステップ５０５で、各生成されたデータセットが、まれな値または一意の値の導入に関して調べられる。ステップ５１０で、各生成されたデータセットが、データ・マスキング・プロセスおよび構成の選択肢の対応するセットに基づいて生じていることがある任意の準識別子を捕捉するために、さらに取り出される。任意の従来またはその他の手法に基づいて、生成されたデータセット内の準識別子が識別されてよい。例えば、生成されたデータセット内の属性のグループによって識別された実体の一意性が、準識別子を決定するために利用されてよく、規則的な表現またはパターンが、既知の準識別子を識別するために使用されてよい、などである。加えて、ユーザは、元のデータ列または統合された列（例えば、同じ種類の列を統合することに基づいて作成された（または適合するテンプレートに従って生成された）列）あるいはその両方から準識別子を指定してよい。

【００５０】

推奨または選択あるいはその両方を決定するために、準識別子の構成要素として識別された、生成されたデータセットの各列が、プライバシーの脆弱性を有しているとしてマーク付けされる。言い換えると、生成されたデータセットを生成するために使用されるデータ・マスキング・プロセスおよび構成の選択肢の対応するセットが、準識別子を生成されたデータセットに導入している。ステップ５１５で、識別された準識別子およびプライバシーの脆弱性が、クライアント・システム１１４上に提示するために提供される。

【００５１】

データのつながりおよび準識別子に関する生成されたデータセットの評価は、任意の順序で実行されてよく、さらに、処理性能を向上させるために、並列に実行されてもよい。加えて、これらの評価の結果は、生成されたデータセット内のプライバシーの脆弱性の存在を決定するための任意の方法で結合されてよい。例えば、特定の数のリンクおよび特定の数の準識別子に応答して、生成されたデータセットに関して、プライバシーの脆弱性が存在することがある。代替として、特定の数のデータ・リンクまたは特定の数の準識別子のいずれかに応答して、プライバシーの脆弱性が存在するということが決定されてよい。この場合、これらの条件のうちの１つが発生したときに、生成されたデータセットがプライ

バシーの脆弱性を有すると見なされ、他の条件に関する追加の処理または評価が終了されてよく、これによって処理時間を短縮する。

【００５２】

データ非特定化プロセスまたは手法で、関連付けられた構成の選択肢の多数のセットを使用してデータセットを生成または評価することは、かなりの処理時間を必要とすることがある。処理性能を向上させ、データを非特定化するための処理時間を短縮するために、本発明の実施形態は、複数の手法を採用してよい。例えば、ユーザによってさまざまなデータ非特定化プロセスおよび構成の選択肢の関連付けられたセットが提供され、評価されてよい。これらのデータ非特定化プロセスの構成のうちの１つまたは複数が、プライバシーの脆弱性を有さないデータセットを生成する場合、残りのデータ非特定化プロセスおよび関連付けられた構成によって生成されるデータセットの生成および評価が終了されてよく、それによって、処理時間を短縮し、計算リソースを保存する。さらに、データ非特定化プロセスが評価する構成の数を示す制限が提供されてよい。

【００５３】

加えて、データ非特定化プロセスおよび構成の選択肢の関連付けられたセットによって生成されるデータセットの生成および評価を制御するために、ツリー構造またはその他のデータ構造が作成されてよく、それによって、計算性能を向上させ、処理時間を短縮する。ツリー構造の形態の例示的なデータ構造が、図８に示されている。例えば、ツリー構造８００は、２つの属性（例えば、名前および住所）の各々に関して２つの構成の選択肢（例えば、削除の選択肢、およびデータ保存の選択肢を使用する非特定化）を含む、非特定化プロセスの構成の選択肢のセットを表している。しかし、ツリー構造は、任意の数の任意の属性に関する任意の非特定化プロセスの任意の数の構成の選択肢を表してよい。

【００５４】

ツリー構造８００は、ルート・ノード８０５ならびにサブツリー８１０および８３０を含んでいる。各ノードは、データ非特定化プロセスの構成の選択肢の対応するセットを表し、対応するテンプレートに関連付けられている。例えば、サブツリー８１０のノード８１２は、第１の属性の構成の選択肢の第１のセット（例えば、名前の削除）を表してよく、一方、ノード８１６は、第１の属性の構成の選択肢の第２のセット（例えば、性別情報を保ちながらの名前の非特定化）を表してよい。ノード８１２の子ノード８１４、８１５は、ノード８１２の構成の選択肢のセット、および第２の属性の構成の選択肢の各セットをそれぞれ表してよい（例えば、名前の削除および住所の削除（ノード８１４）、名前の削除および空間的近接を保ちながらの住所の非特定化（ノード８１５））。ノード８１６の子ノード８１７、８１８は、ノード８１６の構成の選択肢のセット、および第２の属性の構成の選択肢の各セットをそれぞれ表してよい（例えば、性別情報を保ちながらの名前の非特定化および住所の削除（ノード８１７）、性別情報を保ちながらの名前の非特定化および空間的近接を保ちながらの住所の非特定化（ノード８１７））。

【００５５】

同様に、サブツリー８３０のノード８３２は、第２の属性の構成の選択肢の第１のセット（例えば、住所の削除）を表してよく、一方、ノード８３６は、第２の属性の構成の選択肢の第２のセット（例えば、空間的近接を保ちながらの住所の非特定化）を表してよい。ノード８３２の子ノード８３４、８３５は、ノード８３２の構成の選択肢のセット、および第１の属性の構成の選択肢の各セットをそれぞれ表してよい（例えば、住所の削除および名前の削除（ノード８３４）、住所の削除および性別情報を保ちながらの名前の非特定化（ノード８３５））。ノード８３６の子ノード８３７、８３８は、ノード８３６の構成の選択肢のセット、および第１の属性の構成の選択肢の各セットをそれぞれ表してよい（例えば、空間的近接を保ちながらの住所の非特定化および名前の削除（ノード８３７）、空間的近接を保ちながらの住所の非特定化および性別を保ちながらの名前の非特定化（ノード８３８））。重複する（または同じ）構成の選択肢を含むノードが統合されるか、または取り除かれて、各ノードが構成の選択肢の異なるセットを含んでいるツリーを生成する。

【００５６】
ツリー８００内の各親ノードの子ノードは、それらの親ノードと比較してより一般化された情報を含むデータセットを生成する構成の選択肢を表す。例えば、ノード８１２は名前属性を削除してよく、一方、子ノード８１４は名前属性および住所属性の両方を削除してよく、それによって、より詳細でない（または、さらに非特定化された）情報を含むデータセットを生成する。処理中に、ツリー８００がルート・ノード８０５からトラバースされ、宛先ノードの対応するテンプレートがデータ非特定化プロセスに適用されて、データセットを生成する。生成されたデータセットが評価され、プライバシーの脆弱性を有していないということが決定された場合、宛先ノードからの子孫ノードが、より一般化されたデータセットを生成する構成の選択肢に関連付けられているため、これらの子孫ノードも、同様にプライバシーの脆弱性を有していないと見なされる。したがって、評価を実行せずに、子孫ノードがデータ非特定化プロセスの許容できる構成として示され、それによって処理時間を短縮する。
【００５７】
例えば、ノード８１２に対応するテンプレートがデータ非特定化プロセスに適用され、名前属性が削除されたデータセットを生成してよい。このデータセットが評価され、プライバシーの脆弱性を有していないということが決定された場合、名前の削除を超える追加の非特定化を提供するすべての子孫ノード（例えば、ノード８１４、８１５）が、より一般化されたデータを生成するため（例えば、名前の削除および住所の削除（ノード８１４）、名前の削除および住所の非特定化（ノード８１５））、これらの子孫ノードも、プライバシーの脆弱性を有していない。したがって、子孫ノードによって生成されたデータセットを評価するために追加の処理が必要とされず、それによって処理時間を短縮する。
【００５８】
ツリー８００は、データセットの生成または評価あるいはその両方の処理を終了するために使用されてよい。前述したように、親ノードが、最小限のプライバシーの脆弱性を有するか、またはプライバシーの脆弱性を有さないデータセットを生成する構成の選択肢の許容できるセットに関連付けられている場合、子孫ノードに関する処理が終了されてよい。例えば、１つまたは複数のデータ非特定化プロセスのためのデータセットが生成されてよく、処理される生成されたデータセットの数を最小限に抑え、生成されたデータセットの評価時間をより短縮するために、ツリー８００が利用されてよい。この場合、親ノードが、最小限のプライバシーの脆弱性を有するか、またはプライバシーの脆弱性を有さないデータセットを生成する構成の選択肢の許容できるセットに関連付けられている場合、さらに評価を実行せずに、子孫ノードが許容できると見なされる。
【００５９】
代替として、非特定化プロセスで同時に１つまたは複数のノードのデータセットを生成して評価するために、ツリー８００が利用されてよい。これによって、非特定化プロセスが実行されてデータセットを生成する時間を最小限に抑え、評価の数をさらに最小限に抑える。この場合、親ノードが、最小限のプライバシーの脆弱性を有するか、またはプライバシーの脆弱性を有さないデータセットを生成する構成の選択肢の許容できるセットに関連付けられている場合、データセットを生成せず、さらに評価を実行せずに、子孫ノードが許容できると見なされる。
【００６０】
加えて、ツリー８００は、属性のすべてまたは任意の部分に関して、より高いレベルのノードを含むサブツリーを含んでよい。代替として、各サブツリーは、非特定化プロセスの評価のための別個のツリーを形成してよい。
【００６１】
上で説明され、図に示された実施形態が、データ非特定化プロセスの許容できる構成の検出に基づいて、データ非特定化の実施形態を実装する多くの方法のうちのいくつかを表しているにすぎないということが、理解されるであろう。
【００６２】

本発明の実施形態の環境は、任意の望ましい方法で配置された任意の数のコンピュータまたはその他の処理システム（例えば、クライアントまたはエンドユーザ・システム、サーバ・システムなど）およびデータベースまたはその他の保存場所を含んでよく、本発明の実施形態は、任意の望ましい種類のコンピューティング環境（例えば、クラウド・コンピューティング、クライアント／サーバ、ネットワーク・コンピューティング、メインフレーム、スタンドアロン・システムなど）に適用されてよい。本発明の実施形態によって採用されるコンピュータまたはその他の処理システムは、任意の数の任意のパーソナルもしくはその他の種類のコンピュータまたは処理システム（例えば、デスクトップ、ラップトップ、ＰＤＡ、モバイル・デバイスなど）によって実装されてよく、任意の市販のオペレーティング・システムならびに市販のソフトウェアおよびカスタム・ソフトウェア（例えば、ブラウザ・ソフトウェア、通信ソフトウェア、サーバ・ソフトウェア、評価モジュール、データ非特定化プロセスなど）の任意の組み合わせを含んでよい。これらのシステムは、情報の入力または表示あるいはその両方を行うための任意の種類のモニタおよび入力デバイス（例えば、キーボード、マウス、音声認識など）を含んでよい。

【００６３】

本発明の実施形態のソフトウェア（例えば、評価モジュールなど）が、任意の望ましいコンピュータ言語で実装されてよく、本明細書に含まれている機能の説明および図面に示されたフローチャートに基づいて、コンピュータ技術の当業者によって開発され得るということが、理解されるべきである。さらに、本明細書におけるさまざまな機能を実行するソフトウェアのすべての参照は、通常、ソフトウェアの制御下でそれらの機能を実行するコンピュータ・システムまたはプロセッサを参照する。本発明の実施形態のコンピュータ・システムは、代替として、任意の種類のハードウェアまたはその他の処理回路あるいはその両方によって実装されてよい。

【００６４】

コンピュータまたはその他の処理システムのさまざまな機能は、任意の数のソフトウェア・モジュールもしくはユニットまたはハードウェア・モジュールもしくはユニットあるいはその両方、処理システムもしくはコンピュータ・システム、または回路、あるいはその組み合わせの間で、任意の方法で分散されてよく、コンピュータ・システムまたは処理システムは、互いにローカルまたはリモートに配置され、任意の適切な通信媒体（例えば、ＬＡＮ、ＷＡＮ、イントラネット、インターネット、ハードワイヤ、モデム接続、無線など）を介して通信してよい。例えば、本発明の実施形態の機能は、さまざまなエンドユーザ／クライアントおよびサーバ・システム、または任意のその他の中間処理デバイス、あるいはその組み合わせの間で、任意の方法で分散されてよい。上で説明され、フローチャートに示されたソフトウェアまたはアルゴリズムあるいはその両方は、本明細書に記載された機能を実現する任意の方法で、変更されてよい。加えて、フローチャートまたは説明における機能は、望ましい動作を実現する任意の順序で実行されてよい。

【００６５】

本発明の実施形態のソフトウェア（例えば、評価モジュールなど）は、スタンドアロン・システムとネットワークまたはその他の通信媒体によって接続されたシステムと共に使用するための、固定されているか、もしくはポータブルなプログラム製品装置またはデバイスの非一過性のコンピュータ使用可能媒体（例えば、磁気媒体または光媒体、光磁気媒体、フロッピー（Ｒ）・ディスケット、ＣＤ－ＲＯＭ、ＤＶＤ、メモリ・デバイスなど）上で使用可能であってよい。

【００６６】

通信ネットワークは、任意の数の任意の種類の通信ネットワーク（例えば、ＬＡＮ、ＷＡＮ、インターネット、イントラネット、ＶＰＮなど）によって実装されてよい。本発明の実施形態のコンピュータまたはその他の処理システムは、任意の従来またはその他のプロトコルを介してネットワークを経由して通信するために、任意の従来またはその他の通信デバイスを含んでよい。コンピュータまたはその他の処理システムは、ネットワークにアクセスするために、任意の種類の接続（例えば、有線、無線など）を利用してよい。任意

の適切な通信媒体（例えば、ローカル・エリア・ネットワーク（ＬＡＮ）、ハードワイヤ、無線リンク、イントラネットなど）によって、ローカル通信媒体が実装されてよい。

【００６７】

システムは、情報（例えば、元のデータセットおよび暫定的なデータセット、構成または設定、データ非特定化プロセスの選択肢など）を格納するために、任意の数の任意の従来もしくはその他のデータベース、データ・ストア、またはストレージ構造（例えば、ファイル、データベース、データ構造、データ、またはその他の保存場所など）を採用してよい。データベース・システムは、情報を格納するために、任意の数の任意の従来もしくはその他のデータベース、データ・ストア、またはストレージ構造（例えば、ファイル、データベース、データ構造、データ、またはその他の保存場所など）によって実装されてよい。データベース・システムは、サーバ・システムまたはクライアント・システムあるいはその両方に含まれるか、または結合されてよい。データベース・システムまたはストレージ構造あるいはその両方は、コンピュータまたはその他の処理システムからリモートまたはローカルであってよく、任意の望ましいデータを格納してよい。

【００６８】

本発明の実施形態は、情報（例えば、ユーザの嗜好、推奨されるデータ非特定化プロセス、非特定化されたデータセットなど）を取得または提供するために、任意の数の任意の種類のユーザ・インターフェイス（例えば、グラフィカル・ユーザ・インターフェイス（ＧＵＩ：Graphical User Interface）、コマンドライン、プロンプトなど）を採用してよく、このインターフェイスは、任意の方法で配置された任意の情報を含んでよい。このインターフェイスは、任意の適切な入力デバイス（例えば、マウス、キーボードなど）を介して情報を入力／表示し、目的の動作を開始するために、任意の位置に配置された任意の数の任意の種類の入力または作動メカニズム（例えば、ボタン、アイコン、フィールド、ボックス、リンクなど）を含んでよい。このインターフェイスの画面は、任意の方法で画面間を移動するために、任意の適切なアクチュエータ（例えば、リンク、タブなど）を含んでよい。

【００６９】

レポートは、望ましい情報（例えば、推奨、プライバシーの問題など）をユーザに提供するために、任意の方法で配置された任意の情報を含んでよく、ルールまたはその他の基準に基づいて構成可能であってよい。

【００７０】

本発明の実施形態は、前述された特定のタスクまたはアルゴリズムに限定されず、任意の種類の識別子に関して任意のデータ非特定化もしくは匿名化プロセスまたは技術を評価するために利用されてよい。データ非特定化プロセスは、任意の属性を削除または非特定化するための任意の種類の構成の選択肢に関連付けられてよい。構成の選択肢のセットおよびテンプレートは、データ非特定化プロセスの任意の数の任意の構成の選択肢を指定してよい。

【００７１】

生成されたデータセットは、任意の数の任意の種類のプライバシーの脆弱性を識別するために、任意の方法で評価されてよい。生成されたデータセットのデータは、任意の種類の既知またはその他のデータセット（例えば、ユーザによって提供されたデータセット、公開されているデータセット、組織内のデータセットなど）に対してテストされてよい。生成されたデータセットは、任意の数の任意の種類のプライバシーの脆弱性の識別（例えば、任意の数の識別された実体、任意の数の導入された準識別子など）に応答して、脆弱であると見なされてよい。脆弱性を検出するためのしきい値は、任意の望ましい値（例えば、ある数のリンク、ある数の準識別子、ある数のプライバシーの脆弱性など）に設定されてよい。この数は、プライバシーの脆弱性を示すための任意の望ましい方法（例えば、より大きい、より小さい、以上、以下など）で、しきい値と比較されてよい。

【００７２】

構成の選択肢のセット間の関係を識別するために、任意のデータ構造（例えば、ツリー、

階層構造など）が利用されてよい。最小限のプライバシーの脆弱性を有するか、またはプライバシーの脆弱性を有さないデータセットを生成する初期構成に応答して、任意の数の関連する構成の選択肢の処理が終了されてよい。データ非特定化プロセスの構成の選択肢を評価するために、任意の方法でデータ構造がトラバースされてよい。属性のセットに関して、任意の数のデータ非特定化プロセスおよび構成の選択肢の関連付けられたセットが推奨または選択されてよい。例えば、同じまたは異なるデータ非特定化プロセス（および対応する構成）が、データセット内の異なる属性に適用されてよい。

【００７３】

本明細書で使用される用語は、特定の実施形態を説明することのみを目的としており、本発明を制限することを意図していない。本明細書で使用される単数形「ａ」、「ａｎ」、および「ｔｈｅ」は、特に明示的に示されない限り、複数形も含むことが意図されている。「備える」、「備えている」、「含む」、「含んでいる」、「有する」、「有する」、「有している」、「伴う」などの用語は、本明細書で使用される場合、記載された機能、整数、ステップ、動作、要素、またはコンポーネント、あるいはその組み合わせの存在を示すが、１つまたは複数のその他の機能、整数、ステップ、動作、要素、コンポーネント、またはこれらのグループ、あるいはその組み合わせの存在または追加を除外していないということが、さらに理解されるであろう。

【００７４】

下の特許請求の範囲内のすべての手段またはステップおよび機能要素の対応する構造、材料、動作、および等価なものは、具体的に請求されるその他の請求された要素と組み合わせて機能を実行するための任意の構造、材料、または動作を含むことが意図されている。本発明の説明は、例示および説明の目的で提示されているが、網羅的であることは意図されておらず、または開示された形態での発明に制限されない。本発明の範囲および思想を逸脱することなく多くの変更および変形が可能であることは、当業者にとって明らかである。本発明の原理および実際的な適用を最も適切に説明するため、およびその他の当業者が、企図されている特定の用途に適しているようなさまざまな変更を伴う多様な実施形態に関して、本発明を理解できるようにするために、実施形態が選択されて説明された。

【００７５】

本発明のさまざまな実施形態の説明は、例示の目的で提示されているが、網羅的であることは意図されておらず、または開示された実施形態に制限されない。記載された実施形態の範囲および思想を逸脱することなく多くの変更および変形が可能であることは、当業者にとって明らかであろう。本明細書で使用された用語は、実施形態の原理、実際の適用、または市場で見られる技術を超える技術的改良を最も適切に説明するため、あるいは他の当業者が本明細書で開示された実施形態を理解できるようにするために選択された。

【００７６】

本発明は、任意の可能な統合の技術的詳細レベルで、システム、方法、またはコンピュータ・プログラム製品、あるいはその組み合わせであってよい。コンピュータ・プログラム製品は、プロセッサに本発明の態様を実行させるためのコンピュータ可読プログラム命令を含んでいるコンピュータ可読記憶媒体を含んでよい。

【００７７】

コンピュータ可読記憶媒体は、命令実行デバイスによって使用するための命令を保持および格納できる有形のデバイスであることができる。コンピュータ可読記憶媒体は、例えば、電子ストレージ・デバイス、磁気ストレージ・デバイス、光ストレージ・デバイス、電磁ストレージ・デバイス、半導体ストレージ・デバイス、またはこれらの任意の適切な組み合わせであってよいが、これらに限定されない。コンピュータ可読記憶媒体のさらに具体的な例の非網羅的リストは、ポータブル・コンピュータ・ディスケット、ハード・ディスク、ランダム・アクセス・メモリ（ＲＡＭ：random access memory）、読み取り専用メモリ（ＲＯＭ：read-only memory）、消去可能プログラマブル読み取り専用メモリ（ＥＰＲＯＭ：erasable programmable read-only memoryまたはフラッシュ・メモリ）、スタティック・ランダム・アクセス・メモリ（ＳＲＡＭ：static random access

memory）、ポータブル・コンパクト・ディスク読み取り専用メモリ（ＣＤ－ＲＯＭ：compact disc read-only memory）、デジタル多用途ディスク（ＤＶＤ：digital versatile disk）、メモリ・スティック、フロッピー（Ｒ）・ディスク、パンチカードまたは命令が記録されている溝の中の隆起構造などの機械的にエンコードされるデバイス、およびこれらの任意の適切な組み合わせを含む。本明細書において使用されるとき、コンピュータ可読記憶媒体は、それ自体が、電波またはその他の自由に伝搬する電磁波、導波管またはその他の送信媒体を伝搬する電磁波（例えば、光ファイバ・ケーブルを通過する光パルス）、あるいはワイヤを介して送信される電気信号などの一過性の信号自体であると解釈されるべきではない。

【００７８】

本明細書に記載されたコンピュータ可読プログラム命令は、コンピュータ可読記憶媒体から各コンピューティング・デバイス／処理デバイスへ、または、例えばインターネット、ローカル・エリア・ネットワーク、広域ネットワーク、または無線ネットワーク、あるいはその組み合わせといったネットワークを介して外部コンピュータまたは外部ストレージ・デバイスへダウンロードされ得る。このネットワークは、銅伝送ケーブル、光伝送ファイバ、無線送信、ルータ、ファイアウォール、スイッチ、ゲートウェイ・コンピュータ、またはエッジ・サーバ、あるいはその組み合わせを備えてよい。各コンピューティング・デバイス／処理デバイス内のネットワーク・アダプタ・カードまたはネットワーク・インターフェイスは、コンピュータ可読プログラム命令をネットワークから受信し、それらのコンピュータ可読プログラム命令を各コンピューティング・デバイス／処理デバイス内のコンピュータ可読記憶媒体に格納するために転送する。

【００７９】

本発明の動作を実行するためのコンピュータ可読プログラム命令は、アセンブラ命令、命令セット・アーキテクチャ（ＩＳＡ：instruction-set-architecture）命令、マシン命令、マシン依存命令、マイクロコード、ファームウェア命令、状態設定データ、集積回路のための構成データ、あるいは、Ｓｍａｌｌｔａｌｋ（Ｒ）、Ｃ＋＋などのオブジェクト指向プログラミング言語、および「Ｃ」プログラミング言語または同様のプログラミング言語などの手続き型プログラミング言語を含む１つまたは複数のプログラミング言語の任意の組み合わせで記述されたソース・コードまたはオブジェクト・コードであってよい。コンピュータ可読プログラム命令は、ユーザのコンピュータ上で全体的に実行すること、ユーザのコンピュータ上でスタンドアロン・ソフトウェア・パッケージとして部分的に実行すること、ユーザのコンピュータ上で部分的におよびリモート・コンピュータ上でそれぞれ部分的に実行すること、あるいはリモート・コンピュータ上またはサーバ上で全体的に実行することができる。後者のシナリオでは、リモート・コンピュータは、ローカル・エリア・ネットワーク（ＬＡＮ：local area network）または広域ネットワーク（ＷＡＮ：wide area network）を含む任意の種類のネットワークを介してユーザのコンピュータに接続されてよく、または接続は、（例えば、インターネット・サービス・プロバイダを使用してインターネットを介して）外部コンピュータに対して行われてよい。一部の実施形態では、本発明の態様を実行するために、例えばプログラマブル論理回路、フィールドプログラマブル・ゲート・アレイ（ＦＰＧＡ：field-programmable gate arrays）、またはプログラマブル・ロジック・アレイ（ＰＬＡ：programmable logic arrays）を含む電子回路は、コンピュータ可読プログラム命令の状態情報を利用することによって、電子回路をカスタマイズするためのコンピュータ可読プログラム命令を実行してよい。

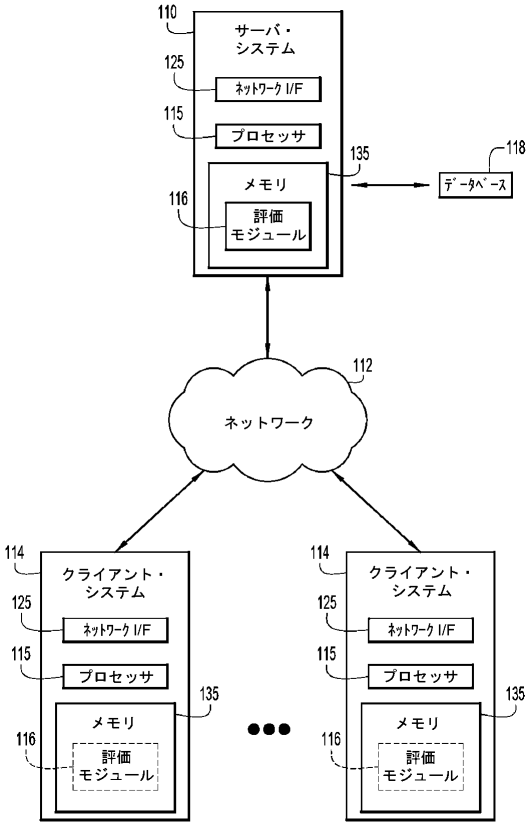
【００８０】

本発明の態様は、本明細書において、本発明の実施形態に従って、方法、装置（システム）、およびコンピュータ・プログラム製品のフローチャート図またはブロック図あるいはその両方を参照して説明される。フローチャート図またはブロック図あるいはその両方の各ブロック、ならびにフローチャート図またはブロック図あるいはその両方に含まれるブロックの組み合わせが、コンピュータ可読プログラム命令によって実装され得るということが理解されるであろう。

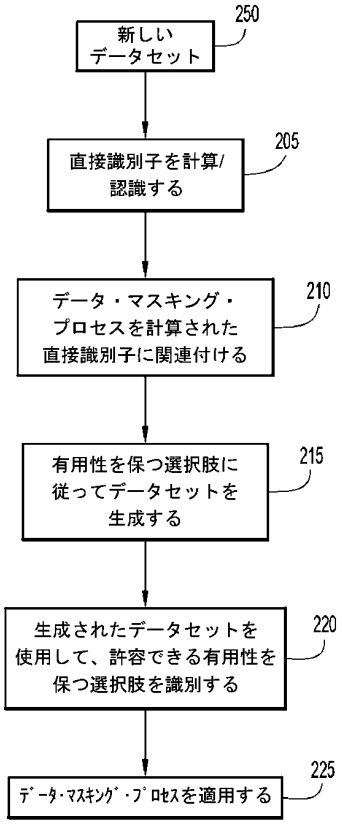
【００８１】
これらのコンピュータ可読プログラム命令は、コンピュータまたはその他のプログラム可能なデータ処理装置のプロセッサを介して実行される命令が、フローチャートまたはブロック図あるいはその両方の１つまたは複数のブロックに指定される機能／動作を実施する手段を作り出すべく、汎用コンピュータ、専用コンピュータ、または他のプログラム可能なデータ処理装置のプロセッサに提供されてマシンを作り出すものであってよい。これらのコンピュータ可読プログラム命令は、命令が格納されたコンピュータ可読記憶媒体がフローチャートまたはブロック図あるいはその両方の１つまたは複数のブロックに指定される機能／動作の態様を実施する命令を含んでいる製品を備えるように、コンピュータ可読記憶媒体に格納され、コンピュータ、プログラム可能なデータ処理装置、または他のデバイス、あるいはその組み合わせに特定の方式で機能するように指示できるものであってもよい。
【００８２】
コンピュータ可読プログラム命令は、コンピュータ上、その他のプログラム可能な装置上、またはその他のデバイス上で実行される命令が、フローチャートまたはブロック図あるいはその両方の１つまたは複数のブロックに指定される機能／動作を実施するように、コンピュータ、その他のプログラム可能なデータ処理装置、またはその他のデバイスに読み込まれてもよく、それによって、一連の動作可能なステップを、コンピュータ上、その他のプログラム可能な装置上、またはコンピュータ実装プロセスを生成するその他のデバイス上で実行させる。
【００８３】
図内のフローチャート図およびブロック図は、本発明のさまざまな実施形態に従って、システム、方法、およびコンピュータ・プログラム製品の可能な実装のアーキテクチャ、機能、および動作を示す。これに関連して、フローチャートまたはブロック図内の各ブロックは、規定された論理機能を実装するための１つまたは複数の実行可能な命令を備える、命令のモジュール、セグメント、または部分を表してよい。一部の代替の実装では、ブロックに示された機能は、図に示された順序とは異なる順序で行われてもよい。例えば、連続して示された２つのブロックは、実際には、含まれている機能に応じて、実質的に同時に実行されるか、または場合によっては逆の順序で実行されてよい。ブロック図またはフローチャート図あるいはその両方の各ブロック、ならびにブロック図またはフローチャート図あるいはその両方に含まれるブロックの組み合わせは、規定された機能または動作を実行するか、あるいは専用ハードウェアとコンピュータ命令の組み合わせを実行する専用ハードウェアベースのシステムによって実装され得るということにも注意する。

【図面】

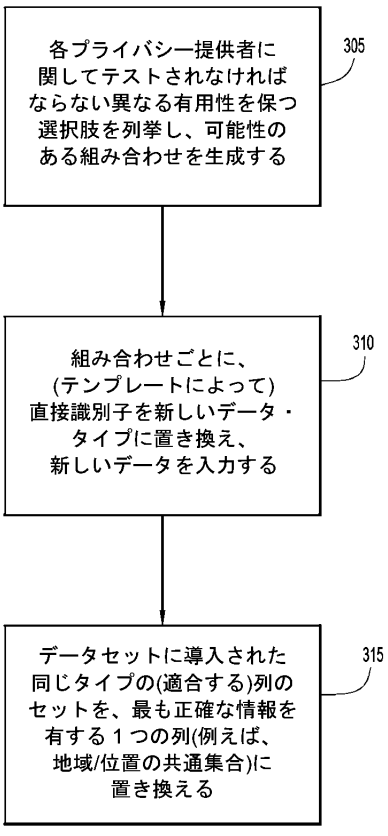
【図１】


110
サーバ・システム
125
ネットワーク I/F
115
プロセッサ
135
メモリ
116
評価モジュール
118
データベース
112
ネットワーク
114
クライアント・システム
125
ネットワーク I/F
115
プロセッサ
135
メモリ
116
評価モジュール
114
クライアント・システム
125
ネットワーク I/F
115
プロセッサ
135
メモリ
116
評価モジュール

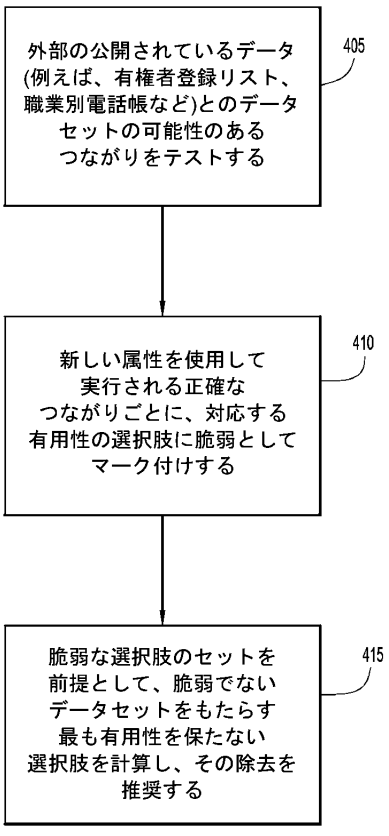
【図２】


250
新しいデータセット
205
直接識別子を計算/認識する
210
データ・マスキング・プロセスを計算された直接識別子に関連付ける
215
有用性を保つ選択肢に従ってデータセットを生成する
220
生成されたデータセットを使用して、許容できる有用性を保つ選択肢を識別する
225
ﾃﾞｰﾀ･ﾏｽｷﾝｸﾞ･ﾌﾟﾛｾｽを適用する

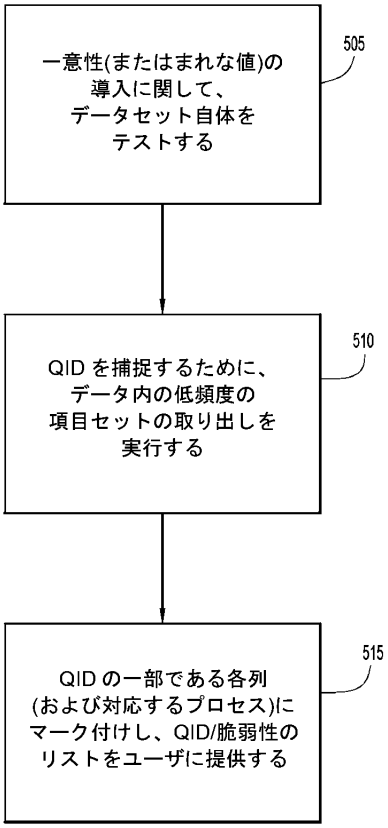
【図３】


各プライバシー提供者に関してテストされなければならない異なる有用性を保つ選択肢を列挙し、可能性のある組み合わせを生成する
305
組み合わせごとに、(テンプレートによって)直接識別子を新しいデータ・タイプに置き換え、新しいデータを入力する
310
データセットに導入された同じタイプの(適合する)列のセットを、最も正確な情報を有する１つの列(例えば、地域/位置の共通集合)に置き換える
315

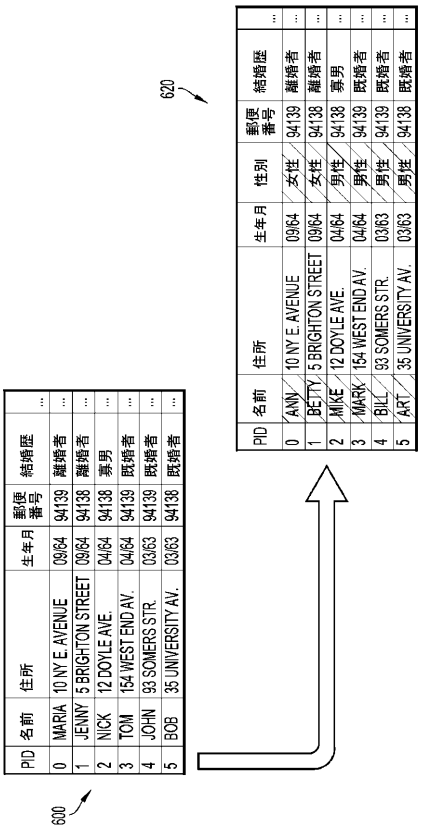
【図４】


外部の公開されているデータ(例えば、有権者登録リスト、職業別電話帳など)とのデータセットの可能性のあるつながりをテストする
405
新しい属性を使用して実行される正確なつながりごとに、対応する有用性の選択肢に脆弱としてマーク付けする
410
脆弱な選択肢のセットを前提として、脆弱でないデータセットをもたらす最も有用性を保たない選択肢を計算し、その除去を推奨する
415

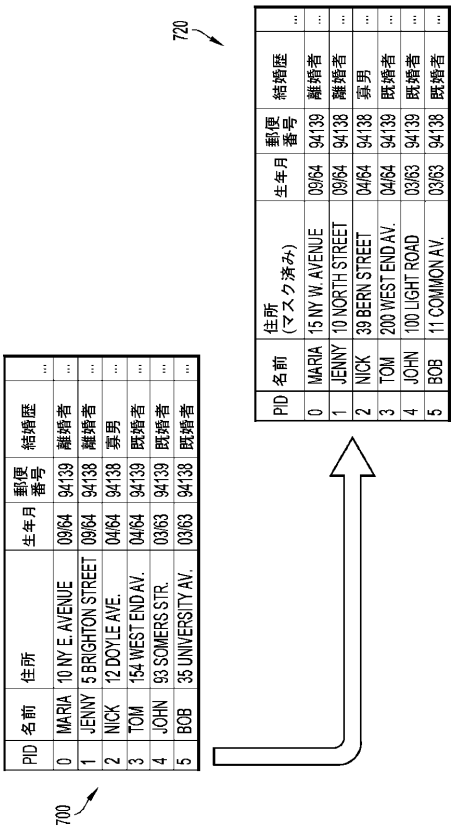
【図５】


一意性(またはまれな値)の導入に関して、データセット自体をテストする
505
QID を捕捉するために、データ内の低頻度の項目セットの取り出しを実行する
510
QID の一部である各列(および対応するプロセス)にマーク付けし、QID/脆弱性のリストをユーザに提供する
515

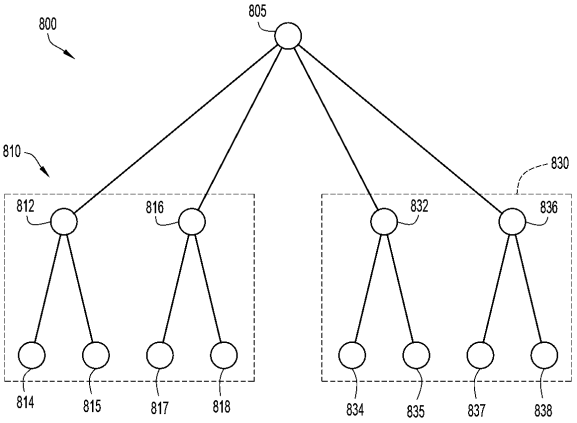
【図６】


620
600
PID 名前 住所 生年月 性別 郵便番号 結婚歴 ...
0 ANN 10 NY E. AVENUE 09/64 女性 94139 離婚者 ...
1 BETTY 5 BRIGHTON STREET 09/64 女性 94138 離婚者 ...
2 MIKE 12 DOYLE AVE. 04/64 男性 94138 寡男 ...
3 MARK 154 WEST END AV 04/64 男性 94139 既婚者 ...
4 BILL 93 SOMERS STR. 03/63 男性 94139 既婚者 ...
5 ART 35 UNIVERSITY AV. 03/63 男性 94138 既婚者 ...
PID 名前 住所 生年月 郵便番号 結婚歴 ...
0 MARIA 10 NY E. AVENUE 09/64 94139 離婚者 ...
1 JENNY 5 BRIGHTON STREET 09/64 94138 離婚者 ...
2 NICK 12 DOYLE AVE. 04/64 94138 寡男 ...
3 TOM 154 WEST END AV. 04/64 94139 既婚者 ...
4 JOHN 93 SOMERS STR. 03/63 94139 既婚者 ...
5 BOB 35 UNIVERSITY AV. 03/63 94138 既婚者 ...

【図７】


720
700
PID 名前 住所 生年月 郵便番号 結婚歴 ...
0 MARIA 10 NY E. AVENUE 09/64 94139 離婚者 ...
1 JENNY 5 BRIGHTON STREET 09/64 94138 離婚者 ...
2 NICK 12 DOYLE AVE. 04/64 94138 寡男 ...
3 TOM 154 WEST END AV. 04/64 94139 既婚者 ...
4 JOHN 93 SOMERS STR. 03/63 94139 既婚者 ...
5 BOB 35 UNIVERSITY AV. 03/63 94138 既婚者 ...
PID 名前 住所(マスク済み) 生年月 郵便番号 結婚歴 ...
0 MARIA 15 NY W. AVENUE 09/64 94139 離婚者 ...
1 JENNY 10 NORTH STREET 09/64 94138 離婚者 ...
2 NICK 39 BERN STREET 04/64 94138 寡男 ...
3 TOM 200 WEST END AV. 04/64 94139 既婚者 ...
4 JOHN 100 LIGHT ROAD 03/63 94139 既婚者 ...
5 BOB 11 COMMON AV. 03/63 94138 既婚者 ...

【図８】


800
805
810
812
816
830
832
836
814
815
817
818
834
835
837
838

フロントページの続き

(72)発明者　グコウララス－ディヴァニス、アリス
アメリカ合衆国０２１４２－１１２３ マサチューセッツ州ケンブリッジ ビニー・ストリート７５

審査官　上島 拓也

(56)参考文献　特開２０１０－０８６１７９（ＪＰ，Ａ）
特開２００８－２１７４２５（ＪＰ，Ａ）
特開２０１７－１７４４５８（ＪＰ，Ａ）
特開２０１７－０４１０４８（ＪＰ，Ａ）

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　２１／６２