(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6752788号
(P6752788)
(45)発行日　令和2年9月9日(2020.9.9)　(24)登録日　令和2年8月21日(2020.8.21)

(51)Int.Cl.　　F I
G06F 16/33 (2019.01)　　G06F 16/33
G06F 21/62 (2013.01)　　G06F 21/62　345

請求項の数 18　(全 22 頁)

(21)出願番号　特願2017-528534(P2017-528534)
(86)(22)出願日　平成27年11月24日(2015.11.24)
(65)公表番号　特表2017-537405(P2017-537405A)
(43)公表日　平成29年12月14日(2017.12.14)
(86)国際出願番号　PCT/US2015/062260
(87)国際公開番号　WO2016/085897
(87)国際公開日　平成28年6月2日(2016.6.2)
審査請求日　平成30年11月26日(2018.11.26)
(31)優先権主張番号　62/084,656
(32)優先日　平成26年11月26日(2014.11.26)
(33)優先権主張国・地域又は機関
米国(US)

(73)特許権者　507061029
レクシスネクシス　ア　ディヴィジョン　オブ　リード　エルザヴィア　インコーポレイテッド
アメリカ合衆国　オハイオ州　４５３４２　マイアミズバーグ　スプリングボロー　パイク　９４４３
(74)代理人　100094569
弁理士　田中　伸一郎
(74)代理人　100088694
弁理士　弟子丸　健
(74)代理人　100103610
弁理士　▲吉▼田　和彦
(74)代理人　100067013
弁理士　大塚　文昭

最終頁に続く

(54)【発明の名称】プライバシーファイアウォールを実装するシステム及び方法

(57)【特許請求の範囲】
【請求項１】
プライベート電子データからプライバシーファイアウォールの外側に位置するリモートコンピュータへ非プライベート情報を決定して提供するプライバシーファイアウォールを実装するシステムであって、
前記プライベート電子データのコーパスを含むデータストレージリポジトリと、
処理装置と、
１又は２以上のプログラム命令を含む非一時的プロセッサ可読記憶媒体と、を備え、
前記１又は２以上のプログラム命令は、実行時に、前記処理装置に、
前記プライバシーファイアウォールの外側に位置する前記リモートコンピュータから、プライベート電子データの前記コーパスの１又は２以上の部分にアクセスするリクエストを含むクエリを受け取ることと、
前記プライバシーファイアウォールの内側のプライベート電子データの前記コーパスを分析して、非プライベート情報を含む第１の１又は２以上のデータ部分と、プライベート情報を含む第２の１又は２以上のデータ部分とを識別することと、
前記第１の１又は２以上のデータ部分に、前記プライバシーファイアウォールの外側で使用できるものとしてタグ付けすることと、
前記第２の１又は２以上のデータ部分が、プライベートではない前記プライベート情報から導出される前記プライベート情報又は追加データの部分である非プライベート要素を含むことを決定することと、

前記第２の１又は２以上のデータ部分から前記非プライベート要素を抽出することであって、前記第２の１又は２以上のデータ部分の残部は非抽出プライベート要素である、抽出することと、
前記非プライベート要素に、前記プライバシーファイアウォールの外側で使用できる情報としてタグ付けすることと、
前記第２の１又は２以上のデータ部分からの前記非抽出プライベート要素を前記プライバシーファイアウォールの外側での使用が拒否されるものとしてタグ付けすることと、
非プライベート情報を含む前記第１の１又は２以上のデータ部分及び前記非プライベート要素のうちの１又は２以上を前記プライバシーファイアウォールの外側に位置する前記リモートコンピュータへ提供することと、
を行わせる、システム。
【請求項２】
実行時に前記処理装置にプライベート電子データの前記コーパスを分析することを行わせる前記１又は２以上のプログラム命令は、前記処理装置に、
プライベート電子データの前記コーパスが除外リスト上のソースから取得されたものであるかどうかを判定することと、
プライベート電子データの前記コーパスが前記除外リスト上のソースから取得されたものである場合、プライベート電子データの前記コーパスに、前記プライバシーファイアウォールの外側での使用が拒否されるものとしてタグ付けすることと、
を更に行わせる、請求項１に記載のシステム。
【請求項３】
実行時に前記処理装置に前記第２の１又は２以上のデータ部分が非プライベート要素を含むかどうかを判定することを行わせる前記１又は２以上のプログラム命令は、前記処理装置に、
前記第２の１又は２以上のデータ部分が、他の場所に出現していた１又は２以上の要素を含むかどうかを判定することと、
前記１又は２以上の下位部分が他の場所に出現していた場合、前記要素に非プライベート要素としてタグ付けすることと、
を更に行わせる、請求項１に記載のシステム。
【請求項４】
実行時に前記処理装置に前記第２の１又は２以上のデータ部分が非プライベート要素を含むかどうかを判定することを行わせる前記１又は２以上のプログラム命令は、前記処理装置に、
前記第２の１又は２以上のデータ部分が、閾値数の異なるソースから出現した１又は２以上の要素を含むかどうかを判定することと、
前記１又は２以上の要素が少なくとも前記閾値数の異なるソースから出現したものである場合、前記１又は２以上の要素に非プライベート要素としてタグ付けすることと、
を更に行わせる、請求項１に記載のシステム。
【請求項５】
実行時に前記処理装置に前記第２の１又は２以上のデータ部分が非プライベート要素を含むかどうかを判定することを行わせる前記１又は２以上のプログラム命令は、前記処理装置に、
前記第２の１又は２以上のデータ部分が、既に公知の情報を含む１又は２以上の要素を含むかどうかを判定することと、
前記１又は２以上の要素が既に公知の情報を含む場合、前記１又は２以上の要素に非プライベート要素としてタグ付けすることと、
を更に行わせる、請求項１に記載のシステム。
【請求項６】
実行時に前記処理装置に前記第２の１又は２以上のデータ部分が非プライベート要素を含むかどうかを判定することを行わせる前記１又は２以上のプログラム命令は、前記処理

装置に、
前記第２の１又は２以上のデータ部分が、１又は２以上の要素が独占情報であると主張できるような複雑な１又は２以上の要素を含むかどうかを判定することと、
前記１又は２以上の要素が複雑である場合、前記１又は２以上の要素に、前記プライバシーファイアウォールの外側での使用が拒否されるものとしてタグ付けすることと、
を更に行わせる、請求項１に記載のシステム。
【請求項７】
実行時に前記処理装置に前記第２の１又は２以上のデータ部分が非プライベート要素を含むかどうかを判定することを行わせる前記１又は２以上のプログラム命令は、前記処理装置に、
前記第２の１又は２以上のデータ部分が、認識可能な連続性を有する１又は２以上の下位部分を含むかどうかを判定することと、
前記１又は２以上の下位部分が認識可能な連続性を有する場合、前記１又は２以上の下位部分に、前記プライバシーファイアウォールの外側での配信が拒否されるものとしてタグ付けすることと、
を更に行わせる、請求項１に記載のシステム。
【請求項８】
実行時に前記処理装置に前記第２の１又は２以上のデータ部分が非プライベート要素を含むかどうかを判定することを行わせる前記１又は２以上のプログラム命令は、前記処理装置に、
前記第２の１又は２以上のデータ部分が、正確なタイムスタンプを有する１又は２以上の要素を含むかどうかを判定することと、
前記１又は２以上の要素が正確なタイムスタンプを有する場合、前記１又は２以上の要素に、前記プライバシーファイアウォールの外側での使用が拒否されるものとしてタグ付けすることと、
を更に行わせる、請求項１に記載のシステム。
【請求項９】
実行時に前記処理装置に前記第２の１又は２以上のデータ部分が非プライベート要素を含むかどうかを判定することを行わせる前記１又は２以上のプログラム命令は、前記処理装置に、
前記第２の１又は２以上のデータ部分が、細分性閾値を下回るタイムスタンプを有する１又は２以上の要素を含むかどうかを判定することと、
前記タイムスタンプが前記細分性閾値を下回る場合、
前記１又は２以上の要素を、前記細分性閾値を上回る時間範囲に調整することと、
前記１又は２以上の調整した要素に非プライベート要素としてタグ付けすることと、
を更に行わせる、請求項１に記載のシステム。
【請求項１０】
実行時に前記処理装置に前記第２の１又は２以上のデータ部分が非プライベート要素を含むかどうかを判定することを行わせる前記１又は２以上のプログラム命令は、前記処理装置に、
前記第２の１又は２以上のデータ部分が、細分性閾値を下回る地理的位置を有する１又は２以上の要素を含むかどうかを判定することと、
前記地理的位置が前記細分性閾値を下回る場合、
前記１又は２以上の要素を、前記細分性閾値を上回る地理的位置に調整することと、
前記１又は２以上の調整した要素に非プライベート要素としてタグ付けすることと、
を更に行わせる、請求項１に記載のシステム。
【請求項１１】
実行時に前記処理装置に前記第１の１又は２以上のデータ部分及び前記非プライベート要素のうちの１又は２以上を前記リモートコンピュータへ提供することを行わせる前記１又は２以上のプログラム命令は、前記処理装置に、

ユーザインターフェイスを介して前記クエリに対する応答をユーザに提供することを更に行わせるものであり、前記クエリに対する前記応答は、非プライベート情報を含む前記第１の１又は２以上のデータ部分及び前記非プライベート要素のうちの１又は２以上から取得された情報を含む、請求項１に記載のシステム。

【請求項１２】

プライベート電子データからプライバシーファイアウォールの外側に位置するリモートコンピュータへ非プライベート情報を決定して提供するプライバシーファイアウォールの実装方法であって、

前記プライバシーファイアウォールの外側に位置する前記リモートコンピュータから、プライベート電子データのコーパスの１又は２以上の部分にアクセスするリクエストを含むクエリを受け取るステップと、

処理装置によって、前記プライバシーファイアウォールの内側のストレージリポジトリに含まれるプライベート電子データの前記コーパスを分析して、非プライベート情報を含む第１の１又は２以上のデータ部分と、プライベート情報を含む第２の１又は２以上のデータ部分とを識別するステップと、

前記処理装置によって、前記第１の１又は２以上のデータ部分に、前記プライバシーファイアウォールの外側で使用できるものとしてタグ付けするステップと、

前記処理装置によって、前記第２の１又は２以上のデータ部分が、プライベートではない前記プライベート情報から導出される前記プライベート情報又は追加データの部分である非プライベート要素を含むかどうかを判定するステップと、

前記処理装置によって、前記第２の１又は２以上のデータ部分から前記非プライベート要素を抽出するステップであって、前記第２の１又は２以上のデータ部分の残部は非抽出プライベート要素である、抽出するステップと、

前記処理装置によって、前記非プライベート要素に、前記プライバシーファイアウォールの外側で使用できる情報としてタグ付けするステップと、

前記処理装置によって、前記第２の１又は２以上のデータ部分からの前記非抽出プライベート要素を前記プライバシーファイアウォールの外側での使用が拒否されるものとしてタグ付けするステップと、

前記処理装置によって、非プライベート情報を含む前記第１の１又は２以上のデータ部分及び前記非プライベート要素のうちの１又は２以上を前記プライバシーファイアウォールの外側に位置する前記リモートコンピュータへ提供するステップと、

を含む方法。

【請求項１３】

プライベート電子データの前記コーパスを分析するステップは、

前記処理装置によって、プライベート電子データの前記コーパスが除外リスト上のソースから取得されたものであるかどうかを判定するステップと、

データの前記コーパスが前記除外リスト上のソースから取得されたものである場合、前記処理装置によって、プライベート電子データの前記コーパスに、前記プライバシーファイアウォールの外側での使用が拒否されるものとしてタグ付けするステップと、

を含む、請求項１２に記載の方法。

【請求項１４】

プライベート電子データからプライバシーファイアウォールの外側に位置するリモートコンピュータへ非プライベート情報を決定して提供するプライバシーファイアウォールの実装方法であって、

処理装置によって、ストレージリポジトリに含まれるプライベート電子データのコーパスを分析して、非プライベート情報を含む第１の１又は２以上のデータ部分と、プライベート情報を含む第２の１又は２以上のデータ部分とを識別するステップと、

前記処理装置によって、前記第１の１又は２以上のデータ部分に、前記プライバシーファイアウォールの外側で使用できるものとしてタグ付けするステップと、

前記処理装置によって、前記第２の１又は２以上のデータ部分が非プライベート要素を

含むかどうかを判定するステップと、
前記第２の１又は２以上のデータ部分が非プライベート要素を含む場合、
前記処理装置によって、前記非プライベート要素を抽出するステップと、
前記処理装置によって、前記非プライベート要素に、前記プライバシーファイアウォールの外側で使用できる情報としてタグ付けするステップと、
前記処理装置によって、非プライベート情報を含む前記第１の１又は２以上のデータ部分及び前記非プライベート要素のうちの１又は２以上を前記リモートコンピュータへ提供するステップと、
を含み、
前記第２の１又は２以上のデータ部分が非プライベート要素を含むかどうかを判定するステップは、
前記処理装置によって、前記第２の１又は２以上のデータ部分が、他の場所に出現していた１又は２以上の要素を含むかどうかを判定するステップと、
前記１又は２以上の下位部分が他の場所に出現していた場合、前記処理装置によって、前記要素に非プライベート要素としてタグ付けするステップと、
前記処理装置によって、前記第２の１又は２以上のデータ部分が、閾値数の異なるソースから出現した１又は２以上の要素を含むかどうかを判定するステップと、
前記１又は２以上の要素が少なくとも前記閾値数の異なるソースから出現したものである場合、前記処理装置によって、前記１又は２以上の要素に非プライベート要素としてタグ付けするステップと、
前記処理装置によって、前記第２の１又は２以上のデータ部分が、既に公知の情報を含む１又は２以上の要素を含むかどうかを判定するステップと、
前記１又は２以上の要素が既に公知の情報を含む場合、前記処理装置によって、前記１又は２以上の下位部分に非プライベート要素としてタグ付けするステップと、
前記処理装置によって、前記第２の１又は２以上のデータ部分が、１又は２以上の要素が独占情報であると主張できるような複雑な１又は２以上の要素を含むかどうかを判定するステップと、
前記１又は２以上の要素が複雑である場合、前記処理装置によって、前記１又は２以上の要素に、前記プライバシーファイアウォールの外側での使用が拒否されるものとしてタグ付けするステップと、
前記処理装置によって、前記第２の１又は２以上のデータ部分が、認識可能な連続性を有する１又は２以上の要素を含むかどうかを判定するステップと、
前記１又は２以上の要素が認識可能な連続性を有する場合、前記処理装置によって、前記１又は２以上の要素に、前記プライバシーファイアウォールの外側での使用が拒否されるものとしてタグ付けするステップと、
前記処理装置によって、前記第２の１又は２以上のデータ部分が、正確なタイムスタンプを有する１又は２以上の要素を含むかどうかを判定するステップと、
前記１又は２以上の要素が正確なタイムスタンプを有する場合、前記処理装置によって、前記１又は２以上の要素に、前記プライバシーファイアウォールの外側での使用が拒否されるものとしてタグ付けするステップと、
を含む方法。
【請求項１５】
前記第２の１又は２以上のデータ部分が非プライベート要素を含むかどうかを判定するステップは、
前記処理装置によって、前記第２の１又は２以上のデータ部分が、細分性閾値を下回るタイムスタンプを有する１又は２以上の要素を含むかどうかを判定するステップと、
前記タイムスタンプが前記細分性閾値を下回る場合、
前記処理装置によって、前記１又は２以上の要素を、前記細分性閾値を上回る時間範囲に調整するステップと、
前記処理装置によって、前記１又は２以上の調整した要素に非プライベート要素とし

てタグ付けするステップと、
を含む、請求項１４に記載の方法。
【請求項１６】
前記第２の１又は２以上のデータ部分が非プライベート要素を含むかどうかを判定するステップは、
前記処理装置によって、前記第２の１又は２以上のデータ部分が、細分性閾値を下回る地理的位置を有する１又は２以上の要素を含むかどうかを判定するステップと、
前記地理的位置が前記細分性閾値を下回る場合、
前記処理装置によって、前記１又は２以上の要素を、前記細分性閾値を上回る地理的位置に調整するステップと、
前記処理装置によって、前記１又は２以上の調整した要素に非プライベート要素としてタグ付けするステップと、
を含む、請求項１４に記載の方法。
【請求項１７】
前記第１の１又は２以上のデータ部分及び前記非プライベート要素のうちの１又は２以上を前記リモートコンピュータへ提供するステップは、
前記処理装置によって、ユーザインターフェイスを介して前記クエリに対する応答をユーザに提供するステップを更に含み、前記クエリに対する前記応答は、非プライベート情報を含む前記第１の１又は２以上のデータ部分及び前記非プライベート要素のうちの１又は２以上から取得された情報を含む、請求項１４に記載の方法。
【請求項１８】
プライバシーファイアウォールを通じてプライベート電子データからプライバシーファイアウォールの外側に位置するリモートコンピュータへ非プライベート情報を提供するシステムであって、
非プライベート情報として、配信すべきでないプライベート情報として、配信すべきプライベート情報の非プライベート要素として、又は非個人化されたプライベート情報としてタグ付けされたプライベート電子データのコーパスを含む、前記プライバシーファイアウォールの内側のデータストレージリポジトリであって、前記非プライベート要素はプライベートではない前記プライベート情報から導出される前記プライベート情報又は追加データの部分である、データストレージリポジトリと、
処理装置と、
１又は２以上のプログラム命令を含む非一時的プロセッサ可読記憶媒体と、を備え、
前記１又は２以上のプログラム命令は、実行時に、前記処理装置に、
前記プライバシーファイアウォールの外側に位置する前記リモートコンピュータから質問を含む検索文字列を受け取ることと、
前記検索文字列に対応する前記プライベート電子データの１又は２以上の部分を求めて前記データストレージリポジトリを検索することと、
前記電子データの前記１又は２以上の部分が、前記非プライベート情報、前記配信すべきプライベート情報の非プライベート要素、又は前記非個人化されたプライベート情報を含む場合、前記リモートコンピュータへ応答を提供することであって、前記応答は、前記非プライベート情報、前記配信すべきプライベート情報の前記非プライベート要素、又は前記非個人化されたプライベート情報を含む前記電子データの前記１又は２以上の部分の内部に格納された情報を含むものである、提供することと、
前記非プライベート情報、前記配信すべきプライベート情報の非プライベート要素、又は前記非個人化されたプライベート情報ではない前記プライベート情報の残部を使用が拒否されるものとしてタグ付けすることと、
を行わせる、システム。
【発明の詳細な説明】
【技術分野】
【０００１】

〔関連出願との相互参照〕
本出願は、２０１４年１１月２６日に出願された「データプライバシーファイアウォールのためのシステム及び方法（ＳＹＳＴＥＭＳ　ＡＮＤ　ＭＥＴＨＯＤＳ　ＦＯＲ　ＤＡＴＡ　ＰＲＩＶＡＣＹ　ＦＩＲＥＷＡＬＬ）」という名称の米国仮特許出願第６２／０８４，６５６号の利益を主張するものであり、この仮特許出願の開示は、全体が引用により本明細書に組み入れられる。

【０００２】
本明細書は、一般にプライベートな機密データを保護するためにプライバシーファイアウォールを提供することに関し、具体的には、プライベートな機密データから取得された非プライベート情報へのアクセスを提供するシステム及び方法に関する。

【背景技術】

【０００３】
現在、プライベートデータ及び／又は機密データへのアクセスは、これらのデータに関連するユーザの行動を匿名化し、この匿名データを閲覧及び使用のために公開することによって提供することができる。しかしながら、このような方法では、データを操作することによってプライベート情報を抽出できるので、プライバシーを効果的に保護できないこともある。また、このような方法が成立しない理由としては、データがもはや有用でなくなる程度にまで匿名化されることも挙げられる。

【発明の概要】

【発明が解決しようとする課題】

【０００４】
したがって、データのプライバシーが侵害されずに、プライベートデータに含まれる情報が有用になるように、プライベートデータを匿名化するのではなくプライベートデータから非プライベート要素を発見して抽出するシステム及び方法が必要とされている。

【課題を解決するための手段】

【０００５】
１つの実施形態では、プライベート電子データから非プライベート情報(non-private information)を決定して提供するプライバシーファイアウォールを実装する(implementing)システムが、プライベート電子データのコーパスを有するデータストレージリポジトリと、処理装置と、非一時的プロセッサ可読記憶媒体とを含む。非一時的プロセッサ可読記憶媒体は、１又は２以上のプログラム命令(programming instructions)を含み、これらの命令は、実行時に、処理装置に、電子データのコーパスを分析して、非プライベート情報を有する第１の１又は２以上のデータ部分と、プライベート情報を有する第２の１又は２以上のデータ部分とを識別することと、第１の１又は２以上のデータ部分に、プライバシーファイアウォールの外側で使用できるものとしてタグ付けすることと、第２の１又は２以上のデータ部分が非プライベート要素(non-private elements)を含むかどうかを判定することと、第２の１又は２以上のデータ部分が非プライベート要素を含む場合、非プライベート要素を抽出し、この非プライベート要素に、プライバシーファイアウォールの外側で使用できる情報としてタグ付けすることと、を行わせる。

【０００６】
別の実施形態では、プライベート電子データから非プライベート情報を決定して提供するプライバシーファイアウォールの実装方法が、処理装置によって、ストレージリポジトリに含まれるプライベート電子データのコーパスを分析して、非プライベート情報を含む第１の１又は２以上のデータ部分と、プライベート情報を含む第２の１又は２以上のデータ部分とを識別するステップと、処理装置によって、第１の１又は２以上のデータ部分に、プライバシーファイアウォールの外側で使用できるものとしてタグ付けするステップと、処理装置によって、第２の１又は２以上のデータ部分が非プライベート要素を含むかどうかを判定するステップと、第２の１又は２以上のデータ部分が非プライベート要素を含む場合、処理装置によって、非プライベート要素を抽出し、処理装置によって、非プライベート要素に、プライバシーファイアウォールの外側で使用できる情報としてタグ付けす

るステップと、を含む。
【０００７】
さらに別の実施形態では、プライバシーファイアウォールを通じてプライベート電子データからの非プライベート情報を提供するシステムが、プライバシーファイアウォールの内側のデータストレージリポジトリと、処理装置と、非一時的プロセッサ可読記憶媒体とを含む。データストレージリポジトリは、非プライベート情報として、配信すべきでないプライベート情報として、配信すべきプライベート情報の非プライベート要素として、又は非個人化されたプライベート情報としてタグ付けされたプライベート電子データのコーパスを含む。非一時的プロセッサ可読記憶媒体は、１又は２以上のプログラム命令を含み、これらの命令は、実行時に、処理装置に、プライバシーファイアウォールの外部のユーザから検索文字列(search string)を受け取ることと、検索文字列に対応する電子データの１又は２以上の部分を求めてデータストレージリポジトリを検索することと、電子データの１又は２以上の部分が、非プライベート情報、配信すべきプライベート情報の非プライベート要素、又は非個人化されたプライベート情報を含む場合、検索文字列に対する応答を提供することと、を行わせる。検索文字列は質問を含み、応答は、電子データの１又は２以上の部分に含まれる、非プライベート情報、配信すべきプライベート情報の非プライベート要素又は非個人化されたプライベート情報を含む情報を含む。
【０００８】
以下の詳細な説明を図面と共に考慮すれば、本明細書で説明する実施形態がもたらすこれらの及びさらなる特徴がさらに十分に理解されるであろう。
【０００９】
図面に示す実施形態は、本質的に説明及び例示のためのものであり、特許請求の範囲に定める主題を限定するものではない。以下の例示的な実施形態についての詳細な説明は、同じ構造を同じ参照数字によって示す以下の図面と共に読むことによって理解することができる。
【図面の簡単な説明】
【００１０】
【図１】本明細書に図示し説明する１又は２以上の実施形態による、プライバシーファイアウォールの内側のデータへのアクセスを提供するシステムの例示的なコンピュータネットワークの概略図である。
【図２】本明細書に図示し説明する１又は２以上の実施形態による、データの提供時に使用できるハードウェア及びソフトウェアをさらに示す図１のサーバコンピュータ装置の概略図である。
【図３】本明細書に図示し説明する１又は２以上の実施形態による、図１のプライバシーファイアウォールの様々なレイヤの概略図である。
【図４】本明細書に図示し説明する１又は２以上の実施形態による、要求に応答してデータを提供する例示的な方法のフロー図である。
【図５】本明細書に図示し説明する１又は２以上の実施形態による、グラフィカルユーザインターフェイスの例示的な検索入力画面の概略図である。
【図６】本明細書に図示し説明する１又は２以上の実施形態による、オートコンプリートオプションを含むグラフィカルユーザインターフェイスの例示的な検索入力画面の概略図である。
【図７】本明細書に図示し説明する１又は２以上の実施形態による、例示的なデータ分析及び分類方法のフロー図である。
【図８】本明細書に図示し説明する１又は２以上の実施形態による、プライベートデータが非プライベート情報を含むかどうかを判定する例示的な方法のフロー図である。
【発明を実施するための形態】
【００１１】
図を全体的に参照すると、本明細書で説明する実施形態は、プライバシーファイアウォールの範囲内に位置するサーバに記憶されたプライベートデータのコーパスへのアクセス

を制限するプライバシーファイアウォールを実装するシステム及び方法に関する。特定の実施形態では、本明細書で説明するシステム及び方法は、一般に１又は２以上のユーザによって提出された質問に応答して、プライベートデータを安全に保ったまま、プライベートデータから取得された情報的回答を提供するように実装することができる。一般に、プライバシーファイアウォールの内側のデータは全てプライベートデータであると想定することができる。しかしながら、実際には、これらのデータの一部が非プライベート情報を含むこともある。また、プライベート情報を含むデータの残部が非プライベート要素を含むこともある。これらの非プライベート情報、及びプライベート情報からの非プライベート要素に、データのプライバシーを維持した状態で、ユーザが提出した質問への回答に使用できるものとしてタグ付けすることができる。

【００１２】

本明細書に開示する方法及びシステムは、例えばデータの使用が「ユーザを横切る」かどうか（すなわち、ある団体のプライベートデータがこの団体以外の誰かによって見られるかどうか、又は推定されるかどうか）に関して不確実性が存在し得る場合、或いは高度に制御されたアクセスを必要としないデータリポジトリを提供することが望ましい場合に使用することができる。プライバシーファイアウォールを適用できる非限定的な例としては、人物Ａから導出されたデータであって、人物Ａにしか影響を与えないと思われる、データリポジトリへのアクセスが厳しく制御されているデータ、人物Ａから導出されたデータであって、人物Ｂに影響を与える可能性がある、又はデータリポジトリへのアクセスが厳しく制御されていないデータ、特定の行動を行ったユーザに（ユーザのＩＤ又はインターネットプロトコル（ＩＰ）アドレスなどを介して）たどり着くことができるデータ、検索文字列保持ポリシーを忠実に守らなければならないデータ（例えば、特定の期間内に削除又は非個人化しなければならないデータ）、プライバシー標準を忠実に守らなければならないデータ、及び軽度に非個人化された（すなわち、ユーザのＩＤ又はＩＰアドレスが削除された）データが挙げられる。

【００１３】

本明細書で使用する「非プライベート情報」という用語は、ある個人又は団体がプライバシーを全く期待しないと思われる情報を含むデータを意味する。非プライベート情報はリポジトリに記憶することができ、最初は、このリポジトリに記憶される全てのデータがプライベートであると想定される。したがって、これらのデータが非プライベート情報を含むことを判別することができる。本明細書で使用する非プライベート情報の説明例は、幅広い異なるソースにわたって非常に一般的に使用されているデータである。非プライベート情報の別の説明例は、特定の個人又は団体に特異的に関連しないデータである。非プライベート情報のさらに別の説明例は、人口統計、データソース、歴史的間隔、地理的範囲及び／又は同様のものなどの、検索文字列から検索を絞り込む構造に関するデータである。非プライベート情報のさらに別の説明例は、インターネット上の誰もが容易に利用できる公開情報、データを含む公的に利用可能な電子フォルダ及び／又は同様のものなどの、非プライベートなウェブ閲覧行動に関するデータである。いくつかの実施形態では、非プライベート情報を非機密データと呼ぶこともできる。いくつかの実施形態では、特定のデータを非プライベートと見なすかどうかを、１又は２以上のルールを適用することによって決定することができる。

【００１４】

本明細書で使用する「プライベートデータ」という用語は、個人又は団体がプライバシーを期待すると思われる情報を含むデータを意味する。プライベートデータの説明例としては、以下に限定するわけではないが、特定の個人又は団体のプライベート情報に関するデータ、（特定のユーザＩＤ、ＩＰアドレス又は同様のものを含むデータなどの）特定の個人又は団体にたどり着くことができるデータ、保持ポリシーの対象であるデータ、（医療保険の携行性と責任に関する法律（ＨＩＰＡＡ）及び／又は同様のものによってプライベートと見なされるデータなどの）特定のプライバシー標準又は規制要件などによってプライベートと見なされるデータ、特定の個人、団体、特定の一群の個人及び／又は団体か

らしか導出できないデータ、独占的なものとして主張される可能性のある複雑なデータ、一般大衆に知られていない情報を含むデータ、及び内部に含まれる情報を誰かが再構築して特定の個人又は団体のプライバシーを侵害する可能性のあるさらなる機密情報を獲得できるデータを挙げることができる。いくつかの実施形態では、特定のデータをプライベートと見なすかどうかを、１又は２以上のルールを適用することによって決定することができる。いくつかの実施形態では、プライバシーファイアウォールの内側のリポジトリに記憶されている全てのデータを、分析されて非プライベート情報が含まれているかどうかが判定されるまで、最初はプライベートデータと見なすことができる。

【００１５】

本明細書で説明するリポジトリ内のデータは、最初はプライベートと見なすことができるが、本明細書では、プライベートデータから導出されるデータ又は追加データの特定の部分をプライベートデータの「非プライベート要素」として分類することができる。非プライベート要素の説明例としては、以下に限定するわけではないが、（プライベート情報が除去されたデータなどの）非個人化されたデータ、（特定数の一意のＩＰアドレスから行われた検索などの）特定数の一意の場所から出現した同一のデータ、特定数の一意の個人及び／又は団体から出現した同一のデータ、プライベートリポジトリに記憶された非プライベート情報、及びプライベートデータから取得される非識別メタデータを挙げることができる。非識別メタデータとしては、以下に限定するわけではないが、プライベートデータを生成した個人又は団体の地理的地域（州、県、区又は地区など、ただしこれ以上具体的ではないもの）、データが生成された日時を示すタイムスタンプ（ただし、分又は秒は除く）、プライベートデータに関連する特定の検索語及び連結子、プライベートデータに関連する市場区分、（特定のウェブブラウザ、検索エンジン又は同様のものなどの）検索に使用された製品、及び検索の結果としての検索結果からのヒット数を挙げることができる。プライベートリポジトリ内に常駐する非プライベート情報識別ルールは、公開されるデータが具体的すぎる場合にはユーザのプライバシーを完全に保護しないことがある。例えば、内部的には一般的な検索文字列が公開されているものの、この文字列が検索を行ったユーザのソースＩＰアドレスも含む場合には、ユーザのプライバシーが侵害される恐れがある。

【００１６】

ユーザデータについては、プライバシーの範囲が存在することができる。例えば、一極では、検索を行っている個人の名前を含む完全な検索文字列をプライベートデータ及び／又は機密データと見なすことができる。他極では、ユーザが検索内のどこかに「ｅ」という文字を使用したというだけの理由によって、他のユーザがいずれかの使用物及び公開物に「ｅ」という文字を使用できなくなるわけではない。これらの両極間には、データがあらゆる種類のプライバシーの含意を有さなくなる中立点が存在することができる。例えば、１０，０００人の異なるユーザが「Ｒｏｅ　ｖ．　Ｗａｄｅ（ロー対ウェイド裁判）」という検索語を入力した後に、続けて米国最高裁判所の判例引用である４１０　Ｕ．Ｓ．　１１３を閲覧した場合、ユーザが「Ｒｏｅ　ｖ．」とタイプし始めた時に直接４１０　Ｕ．Ｓ．　１１３にジャンプすることを提案するユーザインターフェイス（ＵＩ）機能が設けられていれば、たとえこれを行うための情報が特定のユーザの過去の行動に関連していたかもしれない場合であってもプライバシー侵害は存在しない可能性が高い。

【００１７】

ここで図面を参照すると、図１に、本明細書に図示し説明する実施形態による、プライベートデータリポジトリ内の非プライベート情報を決定し、非プライベート情報に基づいて質問への応答を行い、及び／又は非プライベート情報に基づいて検索要求をオートコンプリートするプライバシーファイアウォールを提供するシステムの構成要素を示す例示的なコンピュータネットワークを示している。図１に示すように、コンピュータネットワーク１０は、インターネットなどの広域ネットワーク（ＷＡＮ）、ローカルエリアネットワーク（ＬＡＮ）、モバイル通信ネットワーク、公衆サービス電話ネットワーク（ＰＳＴＮ）、パーソナルエリアネットワーク（ＰＡＮ）、メトロポリタンエリアネットワーク（Ｍ

ＡＮ）、仮想プライベートネットワーク（ＶＰＮ）、及び／又はその他のネットワークを含むことができる。一般に、コンピュータネットワーク１０は、１又は２以上のコンピュータ装置及び／又はこれらの構成要素を電子的に接続するように構成することができる。例示的なコンピュータ装置は、以下に限定するわけではないが、ユーザコンピュータ装置１２ａ、サーバコンピュータ装置１２ｂ及び管理者コンピュータ装置１２ｃを含むことができる。

【００１８】

一般に、ユーザコンピュータ装置１２ａは、ユーザと、コンピュータネットワーク１０に接続された他の構成要素との間のインターフェイスとして使用することができる。したがって、以下でさらに詳細に説明するように、ユーザコンピュータ装置１２ａは、ユーザから１又は２以上の入力を受け取り、又はユーザに情報を提供することなどの、１又は２以上のユーザ対応機能を実行するために使用することができる。また、図１には、管理者コンピュータ装置１２ｃも含まれる。管理者コンピュータ装置１２ｃは、サーバコンピュータ装置１２ｂが監視、更新又は修正を要求した場合に所望の監視、更新及び／又は修正を行うように構成することができる。管理者コンピュータ装置１２ｃは、サーバコンピュータ装置１２ｂに記憶されているコーパスに追加データを入力するために使用することもできる。

【００１９】

サーバコンピュータ装置１２ｂは、１又は２以上のソースからデータを受け取り、データを記憶し、特定のデータ部分から得られる情報へのアクセスが許可されており、情報に配信を許可するタグが付けられている場合には、このような情報を質問への回答又はオートコンプリート提案の形でユーザコンピュータ装置１２ａに提供することができる。一般に、情報を配信できるかどうかの判断は、サーバコンピュータ装置１２ｂとコンピュータネットワーク１０との間に存在するプライバシーファイアウォール１４によって行うことができる。したがって、本明細書でさらに詳細に説明するように、（プライバシーベールと呼ぶこともできる）プライバシーファイアウォール１４は、サーバコンピュータ装置１２ｂに記憶されているデータから得られる特定の情報へのアクセスを許可又は拒否することができる。

【００２０】

なお、ユーザコンピュータ装置１２ａ及び管理者コンピュータ装置１２ｃをパーソナルコンピュータとして示し、サーバコンピュータ装置１２ｂをサーバとして示しているが、これらは非限定的な例であると理解されたい。具体的に言えば、いくつかの実施形態では、これらのいずれかの構成要素には、あらゆるタイプのコンピュータ装置（例えば、モバイルコンピュータ装置、パーソナルコンピュータ、サーバなど）を使用することができる。また、図１にはこれらの各コンピュータ装置を単体のハードウェアとして示しているが、これも一例にすぎない。具体的に言えば、ユーザコンピュータ装置１２ａ、サーバコンピュータ装置１２ｂ及び管理者コンピュータ装置１２ｃの各々は、複数のコンピュータ、サーバ、データベース、構成要素及び／又は同様のものを表すこともできる。

【００２１】

図２に、非プライベート情報を決定し、文書コーパスを検索し、ユーザによって提示された質問への応答を生成し、及び／又はオートコンプリート提案を生成するシステムをさらに示す、図１のサーバコンピュータ装置１２ｂを示す。また、サーバコンピュータ装置１２ｂは、本明細書に図示し説明する実施形態による、ハードウェア、ソフトウェア及び／又はファームウェアとして具体化された、文書コーパスの検索又は検索クエリの生成を行う非一時的コンピュータ可読媒体を含むこともできる。サーバコンピュータ装置１２ｂは、いくつかの実施形態では、必須ハードウェア、ソフトウェア及び／又はファームウェアを含む汎用コンピュータとして構成することができ、いくつかの実施形態では、本明細書で説明する機能を実行するように特異的に設計された専用コンピュータとして構成することもできる。

【００２２】

やはり図２に示すように、サーバコンピュータ装置１２ｂは、プロセッサ３０と、入力／出力ハードウェア３２と、ネットワークインターフェイスハードウェア３４と、（非プライベート情報３８ａ、プライベートデータの非プライベート要素３８ｂ及びその他のデータ３８ｃを記憶することができる）データストレージ要素３６と、非一時的メモリ要素４０とを含むことができる。メモリ要素４０は、揮発性及び／又は不揮発性コンピュータ可読媒体として構成することができ、したがって（ＳＲＡＭ、ＤＲＡＭ及び／又はその他のタイプのランダムアクセスメモリを含む）ランダムアクセスメモリ、フラッシュメモリ、レジスタ、コンパクトディスク（ＣＤ）、デジタル多用途ディスク（ＤＶＤ）及び／又はその他のタイプのストレージ要素を含むことができる。また、メモリ要素４０は、オペレーティングロジック４２及び検索ロジック４４（これらの各々は、一例としてコンピュータプログラム、ファームウェア又はハードウェアとして具体化することができる）を記憶するように構成することもできる。図２には、サーバコンピュータ装置１２ｂの構成要素間の通信を容易にするバス又はその他のインターフェイスとして実装できるローカルインターフェイス４６も含まれる。

【００２３】

プロセッサ３０は、（データストレージ要素３６及び／又はメモリ要素４０などから）命令を受け取って実行するように構成されたいずれかの処理要素を含むことができる。入力／出力ハードウェア３２は、モニタ、キーボード、マウス、プリンタ、カメラ、マイク、スピーカ、タッチ画面、及び／又はデータの受信、送信及び／又は提示を行うその他の装置を含むことができる。ネットワークインターフェイスハードウェア３４は、モデム、ＬＡＮポート、ワイヤレスフィディリティー（Ｗｉ－Ｆｉ）カード、ＷｉＭａｘカード、モバイル通信ハードウェア、及び／又は他のネットワーク及び／又は装置と通信するその他のハードウェアなどの、いずれかの有線又は無線ネットワーキングハードウェアを含むことができる。

【００２４】

なお、データストレージ要素３６は、サーバコンピュータ装置１２ｂの局所及び／又は遠隔地に存在することができ、１又は２以上のデータを記憶して、１又は２以上のデータへのアクセスを選択的に提供するように構成することができると理解されたい。図２に示すように、データストレージ要素３６は、本明細書でさらに詳細に説明するように、非プライベート情報３８ａ、プライベートデータの非プライベート要素３８ｂ、及びその他のデータ３８ｃを記憶することができる。

【００２５】

メモリ要素４０には、オペレーティングロジック４２及び検索ロジック４４が含まれる。オペレーティングロジック４２は、オペレーティングシステム、及び／又はサーバコンピュータ装置１２ｂの構成要素を管理する他のソフトウェアを含むことができる。検索ロジック４４は、以下で詳細に説明するように、グラフィカルユーザインターフェイス内のユーザ入力から検索クエリを生成するように構成することができる。

【００２６】

なお、図２に示す構成要素は例示的なものにすぎず、本開示の範囲を限定するものではないと理解されたい。具体的に言えば、図２の構成要素は、サーバコンピュータ装置１２ｂ内に存在するように示しているが、これは非限定的な例である。いくつかの実施形態では、これらの構成要素のうちの１つ又は２つ以上が、サーバコンピュータ装置１２ｂの外部に存在することもできる。同様に、図２は、サーバコンピュータ装置１２ｂに関するものであるが、ユーザコンピュータ装置１２ａ及び管理者コンピュータ装置１２ｃなどの他の構成要素が同様のハードウェア、ソフトウェア及び／又はファームウェアを含むこともできる。

【００２７】

図３に、図１のプライバシーファイアウォール１４の様々なレイヤを示す。図３に示すレイヤは例示的なものにすぎない。したがって、本開示の範囲から逸脱することなく、これより少ない又はさらなるレイヤを使用することもできる。また、いくつかのレイヤを潰

すことも、又は追加のレイヤにさらに階層化することもできる。各レイヤは、例えばユーザコンピュータ装置１２ａ（図１）のユーザなどの外部要求者に提供される、サーバコンピュータ装置１２ｂ（図１）に含まれるデータへのアクセスの量を表すことができる。アクセスは、一般にデータへの直接アクセスではなく、ユーザによって提出された質問への回答、又はオートコンプリート提案の形を取ることができる。例示的なレイヤは、例えば、分散ソースレイヤ２０、アグリゲーションレイヤ２２、プライバシー施行レイヤ２４及びタスク固有のデータプロバイダレイヤ２６を含むことができる。いくつかの実施形態では、分散ソースレイヤ２０、アグリゲーションレイヤ２２及びプライバシー施行レイヤ２４を、本明細書でさらに詳細に説明するようなデータの調整を行わない限りはこのようなレイヤによって分類されたデータにほとんど又は全くアクセスできない高制約データレイヤとすることができる。いくつかの実施形態では、タスク固有のデータプロバイダレイヤ２６を、このレイヤによって分類されたデータの多く又は全てにアクセスできる低制約データレイヤとすることができる。

【００２８】

分散ソースレイヤ２０は、例えば、顧客セッション行動ソース内で典型的に見られるデータを分類することができる。このような顧客セッション行動ソースは、受け取って記憶したデータが元々存在していた複数の異なるプラットフォーム上及び／又はアプリケーション上に存在する複数のソースを表すことができる。非限定的な例として、１つのソースは、特定のプログラム又はアプリケーションの検索ボックスとすることができる。異なるソースからのデータは、異なるデータリポジトリに記憶することができる。いくつかのデータリポジトリは、他のデータリポジトリよりも多くのデータ制約を有することができる。したがって、データは、異なるリポジトリにわたって正規化されないこともある。

【００２９】

アグリゲーションレイヤ２２は、例えば、正規化データリポジトリ内で典型的に見られるデータを分類することができる。すなわち、これらのデータは、様々なデータリポジトリから各々をそれぞれのネイティブフォーマットで取得し、１又は２以上の正規化ツール（「ノーマライザ」）によって単一の一貫したフォーマットに正規化しておくことができる。いくつかの実施形態では、データを正規化すると、データに含まれる機密情報の量が最小になるようにできるだけ実用的に匿名化することができる。しかしながら、いくつかの実施形態では、リポジトリが、特定のモジュールにとって十分な一連の共通する一貫した属性を含むことがあり、したがって特定のクエリ（すなわち、そのデータに関連するクエリ）が提示されるまで完全な匿名化が不可能な場合もある。したがって、匿名化に関わらず、リポジトリに含まれるデータは、依然として機密性の高いものとなり得る。したがって、このようなデータへのアクセスは、高度に制限することができる。

【００３０】

プライバシー施行レイヤ２４は、タスク固有のレポータ及びサニタイザモジュールを通過したデータを分類することができる。このようなデータは、例えば、特定の方法でクエリが提示された場合に機密情報を含むことができる正規化データとすることができる。例えば、複数の異なる方法で複数の質問を行い、ブール代数演算を用いて公開予定よりも多くのデータを引き出すことによって、プライバシーベールに穴を開けることができる。したがって、特定の情報を取得するために行うことができる質問のタイプを厳しく制限し、公開される全ての様々な質問の回答が組み合わされることによって情報が漏れることがないように保証することが望ましいと考えられる。したがって、プライバシーファイアウォール１４に提示される各クエリは、結果を配信するように形成された特定のモジュールを有することができる。各特定のモジュールは、機密データを非機密データに変換するブリッジとして機能するように、高度に制限された環境で厳しく吟味して構築することができる。

【００３１】

一般に、タスク固有のデータプロバイダレイヤ２６は、非プライベートな脱感作したプライベートデータ、又はプライベートデータの非プライベート要素である情報のタスク固

有の公的リポジトリを含むことができる。このようなデータは、クエリへの回答に使用することができる。

【００３２】

次に、図４に、１又は２以上の実施形態によるプライバシーファイアウォールの実装方法のフロー図を示す。本明細書でさらに詳細に説明するように、実施形態では、ユーザが情報を要求し、非プライベート情報及び／又はプライベートデータの非プライベート要素を見ることができる。ブロック１８０において、システムは、ユーザコンピュータ装置１２ａのディスプレイ装置上に表示するグラフィカルユーザインターフェイスを生成することができる。このグラフィカルユーザインターフェイスは、ブロック１８２においてユーザが検索文字列を送信できるように構成される。図５を参照すると、グラフィカルユーザインターフェイスは、ユーザからの検索文字列を受け取るように構成された検索文字列入力画面１００を含むことができる。なお、実施形態は、図全体を通じて示すグラフィカルユーザインターフェイスの構成に限定されるものではなく、他のグラフィカルユーザインターフェイス構成も可能であると理解されたい。１つの実施形態では、ネットワーク１０がインターネットであり、本明細書で説明するグラフィカルユーザインターフェイスが、ウェブブラウザを介してユーザに提示される。

【００３３】

検索文字列入力画面１００は、検索文字列を構成する１又は２以上の用語をユーザが（キーボードを用いて）入力できる検索文字列フィールド１０３を含む。１つの実施形態では、検索文字列を、自然言語の検索文字列とすることができる。例えば、ユーザは、例えば「妊娠中絶に関する最高裁判所の画期的な裁判は？」などの質問を行うことができる。別の例では、図５に示す実施形態のように、特定のユーザが１９７３年の米国最高裁判決に関する情報を検索することに関心を持っている可能性がある時に、検索文字列フィールド１０３に「Ｒｏｅ　ｖ．　Ｗａｄｅ（ロー対ウェイド裁判）」という検索語が入力されている。いくつかの実施形態では、図６に示すように、ユーザが検索文字列フィールド１０３に１又は２以上の文字を入力するだけでよく、システムは、サーバコンピュータ装置１２ｂに含まれるプライベートデータリポジトリから取得されたデータに基づいて、提案するオートコンプリートオプションを生成することができる。例えば、ユーザが、判例データベースを検索している場合に「ＲＯ」という文字をタイプすると、検索文字列フィールド１０４内のユーザ入力の下方に示すような、例えばＲｏｅ　ｖ．　Ｗａｄｅ（ロー対ウェイド裁判）、Ｉｎ　Ｒｅ　Ｒｏｓｓ（Ｒｏｓｓについて）及び／又は同様のものなどの、「ＲＯ」という文字を含むいくつかのオートコンプリートオプションをユーザに提示することができる。オートコンプリート提案は、例えば多くのユーザがこれらの特定の裁判事件を検索したことをデータが示す結果として生成することができ、任意に注目度順にランク付けすることができる。したがって、システムは、サーバコンピュータ装置１２ｂに含まれるプライベートデータに基づいて、ユーザが検索文字列をタイプし終える前にユーザが何を検索したいと思っているかを推測しようと試みることができる。しかしながら、オートコンプリート提案を行うために取得する情報は、一般に非プライベート情報及び／又はプライベートデータの非プライベート要素とすることができるので、プライバシーを侵害する可能性のある用語はオートコンプリート提案に含まれない。

【００３４】

図５及び図６に示すように、検索文字列入力画面１００は、オプションボタン１０６、セクションフィルタ入力部１０２及び検索開始アイコン１０５などの他の入力機能を含むこともできる。なお、これより多くの又は少ない入力機能を使用することもできると理解されたい。図５及び図６に示す例では、オプションボタン１０６を使用すると、ユーザは、検索中のコーパスのネイティブ言語ではない電子データの機械翻訳を検索することもできる。他のオプションを提供することもできる。セクションフィルタ入力部１０２を使用すると、ユーザは、特定のデータセクション又はデータセクションの組み合わせのみを検索することができる。例えば、訴訟の事例では、ユーザは、セクションフィルタ入力部１０２を使用して、訴訟概要セクション、キーワードセクション、裁判所の見解セクション

、事実セクション及び／又は同様のもののみを検索することができる。
【００３５】
ユーザは、検索開始アイコン１０５をクリック又は別様に選択することにより、検索文字列フィールド１０３に入力された検索文字列に基づく検索を開始することができる。ブロック１８４において、検索文字列の個々の用語をクエリ語として使用して検索文字列を分析する。一般に、検索文字列の分析は、ユーザが何を検索しているかを特定することを含むことができ、この特定は、現在知られている又は今後構築されるいずれかの方法によって行うことができる。いくつかの実施形態では、ブロック１８６において、ユーザが提出した検索文字列に基づいて適切な検索クエリを決定することができる。すなわち、ユーザによって提出された検索文字列を受け取って解釈し、アクセスすべきデータ、データを含むシステムのタイプ及び／又は同様のものに基づいて適切な検索クエリを生成することができる。適切な検索クエリは、あらゆる数のクエリ生成技術を用いて生成することができる。例えば、ユーザによって提出された検索文字列の分析に基づいて、ブール加重検索クエリを生成することができる。
【００３６】
ブロック１８８において、ファイアウォール１４の内側のサーバコンピュータ装置１２ｂなどのリポジトリに含まれるデータを検索することができる。検索クエリに関連するデータが見つかると、このデータを分析し、ステップ１９０において、データがプライベートであるか、それとも非プライベートであるかを判定することができる。いくつかの実施形態では、プライベートデータをさらに分析し、ステップ１９２において、このデータが非プライベート要素を含むかどうかを判定することができる。また、プライベートデータを分析し、ステップ１９４において、いくつかの要素を非プライベート要素になるように調整できるかどうかを判定することもできる。データを調整できる場合、ステップ１９６においてこのような調整を行い、（非個人化データとも呼ばれる）調整済みデータを取得することができる。例えば、これらのデータは、日時スタンプの細分性の加減、地理的位置の細分性の加減、及び／又は同様のものを行うように調整することができる。いくつかの実施形態では、特定の細分性閾値(granularity threshold)を上回るように細分性を加減することができる。例えば、州又は県と市との間を地理的細分性閾値とすることができる（例えば、州、県、国、区などの細分性は閾値を「上回る」ものとし、市、所在地住所又は同様のものなどのさらなる詳細は閾値を「下回る」ものとすることができる）。時間細分性閾値は、例えば時間と分との間とすることができる（例えば、時、日、週、月及び年で表示された時間は閾値を「上回る」ものとし、分及び秒で表示された時間は閾値を「下回る」ものとすることができる）。いくつかの実施形態では、ステップ１９０、１９２、１９４及び１９６を検索クエリ毎に行うことができる。他の実施形態では、ステップ１８８における検索を素早く行えるように、リポジトリ内にデータが取得されると直ちにステップ１９０、１９２、１９４及び１９６を行うことができる。
【００３７】
ステップ１９８において、検索クエリのクエリ語を用いてデータのコーパスを検索し、検索文字列への応答を行う。一般に、この応答は、非プライベート情報から、及び／又はプライベートデータの調整部分（存在する場合）を含むプライベートデータの非プライベート要素から導出される。応答は、ユーザが行った質問に応答する自然言語の回答、特定の参照への１又は２以上のリンク、戻されて表示される一連の電子文書、及び／又は同様のものとすることができる。
【００３８】
再び図１を参照すると、様々な実施形態では、新たなデータが生成され及び／又は利用可能になると、サーバコンピュータ装置１２ｂが、これらのデータを収集し続けることができる。これらの新たなデータを分析して、プライベートデータ又は機密データを含むかどうかについての判定を行うことができ、本明細書で説明するように、プライベートデータ又は機密データはファイアウォール１４を通り抜けることができない。したがって、図７に、データ分析及び分類処理を示す。

【００３９】
図７に示すように、また図１も参照すると、ステップ２０２において、リポジトリからデータを取得することができる。例えば、いくつかの実施形態では、リモートリポジトリからファイアウォール１４の内側のサーバコンピュータ装置１２ｂにデータをコピー又は移動することができる。他の実施形態では、サーバコンピュータ装置１２ｂ内（例えば、データストレージ３６（図２）内）又はファイアウォール１４の内側の別の場所にリポジトリを配置することにより、コピー又は転送を不要にすることができる。
【００４０】
本明細書で説明したように、最初はこれらのデータをプライベートデータと見なすことができる。ステップ２０４においてデータを分析し、ステップ２０６において、データが非プライベート情報を含むかどうかを判定することができる。データが、プライベートな可能性のある部分を全く含まない場合には、このデータを非プライベート情報として識別することができ、ステップ２０８において、プライバシーファイアウォール１４の外側で使用できるものとしてタグ付けすることができる。例えば、データが、公開裁判記録、公的に入手可能な不動産記録及び／又は同様のものなどの公的に入手可能な情報を全体的に含む場合、ステップ２０６において、データがプライベートデータを含んでいないと判定し、ステップ２０８において、プライバシーファイアウォール１４の外側で使用できるものとしてタグ付けすることができる。
【００４１】
一方、データのいずれかの部分が、機密的な、プライベートな、或いは機密的又はプライベートと見なすことができる情報を含む場合、このデータは、引き続きプライベートデータとして識別することができる。例えば、データが裁判記録を含んでいるが、これらの裁判記録が個人の住所などのプライベート情報を含む場合、このデータはプライベートデータとして識別することができる。
【００４２】
図８に示すように、データが非プライベート情報を含むかどうかの判定は、複数の判定ステップを含むことができる。このようなステップは例示的なものにすぎず、本開示の範囲から逸脱することなく別のステップ、さらなるステップ又はこれより少ないステップを行うこともできると理解されたい。さらに、実施形態は、この図８に示すステップ順によって限定されるものではないと理解されたい。ステップ２０６ａに示すような１つの例示的なステップでは、データが除外リスト上のソースから収集されたものであるかどうかを判定することができる。例示的な除外リストとしては、例えば、ＨＩＰＡＡ、医療相互運用性試験及び適合度（ＨＩＴＣＨ）プロジェクト、ドライバー個人情報保護法（ＤＰＰＡ）、（１９９９年金融サービス現代化法としても知られている）グラム・リーチ・ブライリー法（ＧＬＢＡ）、ＰＣＩデータセキュリティ基準（ＰＣＩ　ＤＳＳ）、及び／又は同様のものによってプライベートとして指定されている情報を含むリストを挙げることができる。データが除外リスト上のソースから収集されたものである場合、ステップ２０９において、このデータをプライベートとして識別することができる。データが除外リスト上のソースから収集されたものでない場合には、さらなる判定ステップを行うことができ、或いはステップ２０７において、非プライベート情報を含むものとしてデータを識別し、ステップ２０８において、プライバシーファイアウォールの外側で使用できるものとしてタグ付けすることができる。
【００４３】
例示的なステップ２０６ｂでは、同じデータが他の場所に出現したことがあるかどうかを判定することができる。例えば、検索エンジンインターフェイスに入力された特定の検索文字列に関するデータを収集する場合、ステップ２０６ｂにおける判定は、別の場所から同じ検索文字列が入力されたことがあるかどうかを確認しようと試みることができる。これらの場所は、メタデータ又は同様のものを再検討することによって判定することができる。例えば、特定の検索クエリに関するメタデータは、検索エンジンインターフェイスに検索文字列を入力するために使用した装置のＩＰアドレスを含むことができる。同じ検

索クエリが異なるＩＰアドレスによって入力され、このようなＩＰアドレスが異なる場所を構成している（すなわち、同じ物理的位置に由来するＩＰアドレスでない）場合には、このデータが他の場所に出現したことがあると判定することができる。データが他の場所に出現したことがない場合、ステップ２０９において、このデータをプライベートとして識別することができる。データが他の場所に出現したことがある場合には、さらなる判定ステップを行うことができ、或いはステップ２０７において、非プライベート情報を含むものとしてデータを識別し、ステップ２０８において、プライバシーファイアウォールの外側で使用できるものとしてタグ付けすることができる。

【００４４】

例示的なステップ２０６ｃでは、データが少なくとも２０個の異なるソースから取得されたものであるかどうかを判定することができる。例えば、データが、データのソースを示すメタデータ（例えば、ＩＰアドレス又は同様のもの）を含む場合、この判定は、メタデータを再検討して、データが２０個の異なるソースから取得されたものであることを保証するステップを含むことができる。なお、本明細書で使用するソースの数（２０）は例示的なものにすぎず、あらゆる数のソース、特にデータがプライベートではないことを保証する数のソースを指定することができると理解されたい。例えば、ソースの数は、データのタイプ、データに関する特定のルール又はポリシー及び／又は同様のものに基づいて様々とすることができる。データが少なくとも２０個の異なるソースから出現したものでない場合、ステップ２０９において、このデータをプライベートとして識別することができる。データが少なくとも２０個の異なるソースから出現したものである場合には、さらなる判定ステップを行うことができ、或いはステップ２０７において、非プライベート情報を含むものとしてデータを識別し、ステップ２０８において、プライバシーファイアウォールの外側で使用できるものとしてタグ付けすることができる。

【００４５】

例示的なステップ２０６ｄでは、データが既に公知の情報を含むかどうかを判定することができる。例えば、データが、通常はプライベートと思われる個人の住所などの情報を含むが、その個人が自宅外で営む事業を宣伝するために自身の自宅住所を公表していた場合、このような情報は、既に公知と見なすことができる。データが、未だ公知ではない情報を含む場合、ステップ２０９において、このデータをプライベートとして識別することができる。データが、既に公知の情報を含む場合には、さらなる判定ステップを行うことができ、或いはステップ２０７において、非プライベート情報を含むものとしてデータを識別し、ステップ２０８において、プライバシーファイアウォールの外側で使用できるものとしてタグ付けすることができる。

【００４６】

例示的なステップ２０６ｅでは、データが独占的であると主張されるほど十分に複雑なものであるかどうかを判定することができる。データの複雑性は、データの性質、データソースの性質、データが収集された状況、及びデータの提供者との間で結ばれたいずれかの合意（例えば、利用合意条件）に基づく個々の場合に応じて決定することができる。例えば、ある団体が独占的であると主張できる（例えば、企業秘密などの）複雑なアルゴリズムを開発して、このアルゴリズムがデータ内に出現している場合、少なくともこのアルゴリズムを含むデータ部分はプライベートとして見なすことができる。別の例では、ある団体が、非常に解釈の狭い及び／又は一意の検索文字列を入力した場合、この検索文字列は複雑なものと見なすことができる。したがって、データが、独占的であると主張されるほど十分に複雑なものであると判定された場合、ステップ２０９において、このデータをプライベートデータとして識別することができる。データが、独占的であると主張されるほど十分に複雑なものでない場合には、さらなる判定ステップを行うことができ、或いはステップ２０７において、非プライベート情報を含むものとしてデータを識別し、ステップ２０８において、プライバシーファイアウォールの外側で使用できるものとしてタグ付けすることができる。

【００４７】

例示的なステップ２０６ｆでは、データが正確なタイムスタンプを含むかどうかを判定することができる。例えば、データが、検索を行った個人又は団体を他者が識別できる正確なタイムスタンプを有する検索文字列に関連する場合、このデータはプライベートとすることができる。本明細書で説明したようなタイムスタンプの細分性を加減するように（例えば、検索文字列が入力された時間よりも具体的でないように）データが適切に調整されない限り、ステップ２０９において、このデータをプライベートとして識別することができる。データが正確なタイムスタンプを含まない又はデータが適切に調整されていた場合には、さらなる判定ステップを行うことができ、或いはステップ２０７において、非プライベート情報を含むものとしてデータを識別し、ステップ２０８において、プライバシーファイアウォールの外側で使用できるものとしてタグ付けすることができる。

【００４８】

例示的なステップ２０６ｇでは、データが認識可能な連続性を含むかどうかを判定することができる。一般に、認識可能な連続性とは、たとえ単独のデータ文字列はプライベートとして見なされないと思われる場合でも、合わせて見た時にはプライベートに違いない情報を含むデータ文字列の連続性とすることができる。例えば、データが、そのデータに関連する個々の団体又はデータを特定できる可能性のある連続性を含む場合、ステップ２０９において、このようなデータをプライベートデータとして識別することができる。別の例では、ある団体が、「製造に基づく集団訴訟に適した裁判所」、「スクレーパウィジェットの製造会社」、「コネチカット州の会社」及び「切り傷でコネチカット州の救急治療室に搬送された個人」という用語を用いて４回の後続する検索を行った場合、この団体は、コネチカット州のスクレーパウィジェットメーカーに対して集団訴訟の申し立てを検討している代理人又は法律事務所であると推測することができ、たとえこれらの検索文字列を個別に見た時にはこのような情報が明らかになることはなくプライベートな可能性はない場合でも、プライベート情報を構成する可能性がある。データが、そのデータに関連する個人又は団体を識別するために使用できる連続性を含んでいない場合には、さらなる判定ステップを行うことができ、或いはステップ２０７において、非プライベート情報を含むものとしてデータを識別し、ステップ２０８において、プライバシーファイアウォールの外側で使用できるものとしてタグ付けすることができる。

【００４９】

再び図７を参照すると、ステップ２１０において、プライベートデータが非プライベート要素を含むかどうかを判定することができる。すなわち、プライベートデータが、単独時に非プライベート情報を構成する特定の部分を有する場合、このようなプライベートデータは非プライベート部分を含むものとして識別され、ステップ２１４において、このデータから非プライベート要素を抽出することができる。本明細書で上述したように、非プライベート要素は、非プライベート要素からプライベート情報が収集されないと考えられる形で抽出することができる。例えば、非プライベート要素が、プライベートデータから抽出できるメタデータである場合、このようなメタデータは、範囲が限定されたものでなければならない（例えば、ユーザの地理的位置が、ユーザが存在する州、地区、県などよりも具体的にならないようにすることができる）。非プライベート要素は、抽出されると、ステップ２１６において、プライバシーファイアウォールの外側で使用できるものとしてタグ付けすることができる。データが非プライベート要素を含まない場合、このデータには、ステップ２１２において、プライバシーファイアウォール１４を通じてアクセスできない拒否データとしてタグ付けすることができる。

【００５０】

なお、本明細書で説明した実施形態は、データを生成した個人又は団体のプライバシーを侵害することなくプライベートデータへのアクセスを可能にするプライバシーファイアウォールを提供するシステム及び方法を提供するものであると理解されたい。本明細書のシステム及び方法の本質は、回答を得るためにプライベートデータへのアクセスを必要とする可能性がある質問又は検索文字列の入力をユーザが行えるようにするものである。この時、ユーザは、プライベートデータリポジトリへのアクセス権を得ることなく、質問に

対する応答又はオートコンプリート提案を受け取ることができる。

【００５１】

本明細書では特定の実施形態を図示し説明したが、特許請求する主題の趣旨及び範囲から逸脱することなく、他の様々な変更及び修正を行うことができると理解されたい。さらに、本明細書では、特許請求する主題の様々な態様について説明したが、このような態様を組み合わせて利用する必要はない。したがって、添付の特許請求の範囲は、特許請求する主題の範囲に含まれるこのような全ての変更及び修正を対象にすることが意図されている。

【図１】

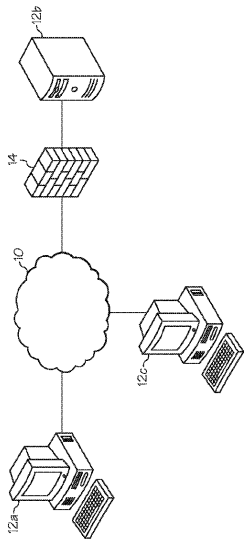

FIG. 1

【図２】

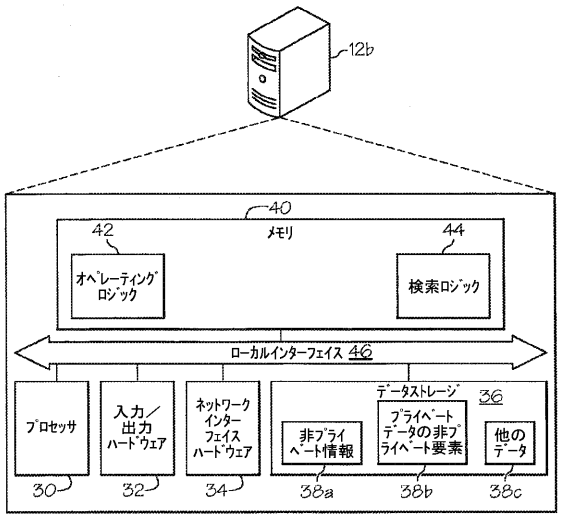

FIG. 2

【図３】

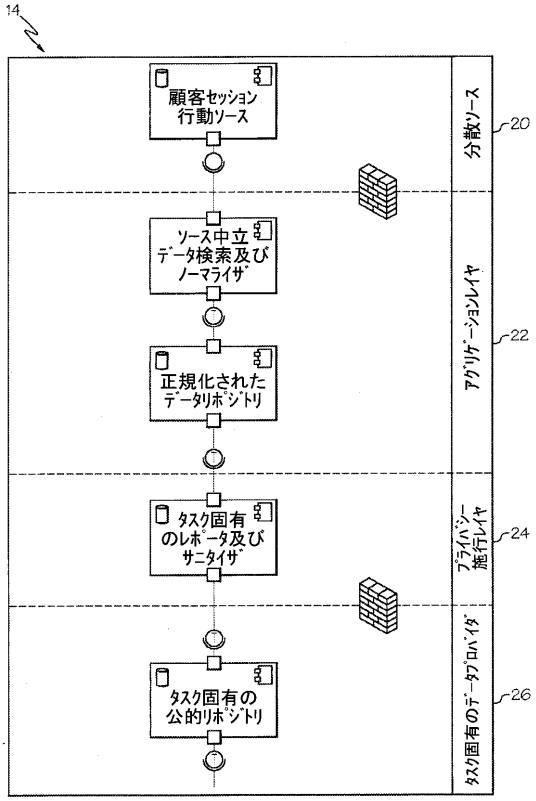

FIG. 3

【図４】

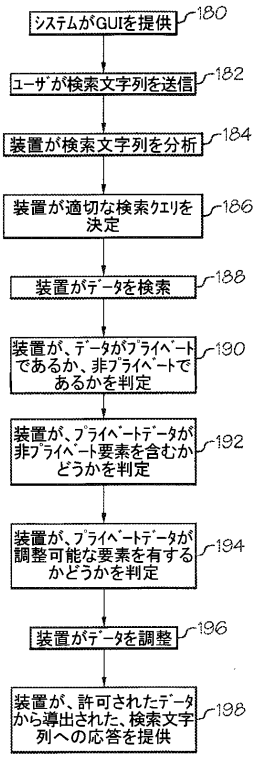

FIG. 4

【図５】

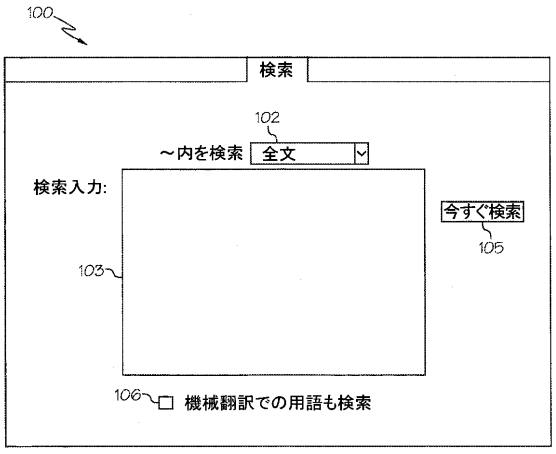

FIG. 5

【図６】

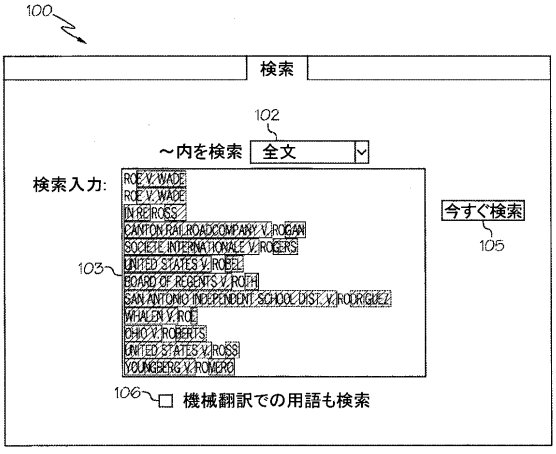

FIG. 6

【図７】

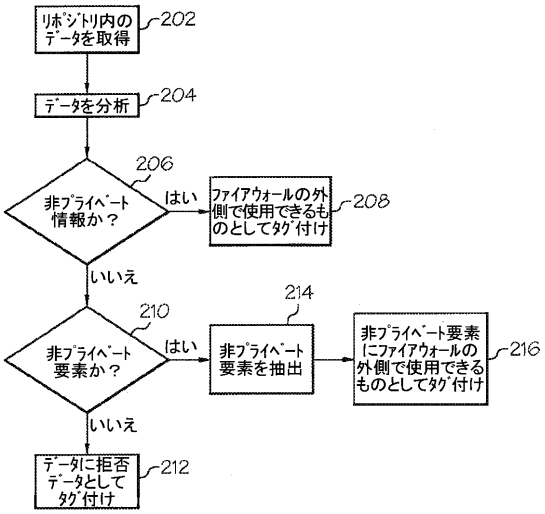

FIG. 7

【図８】

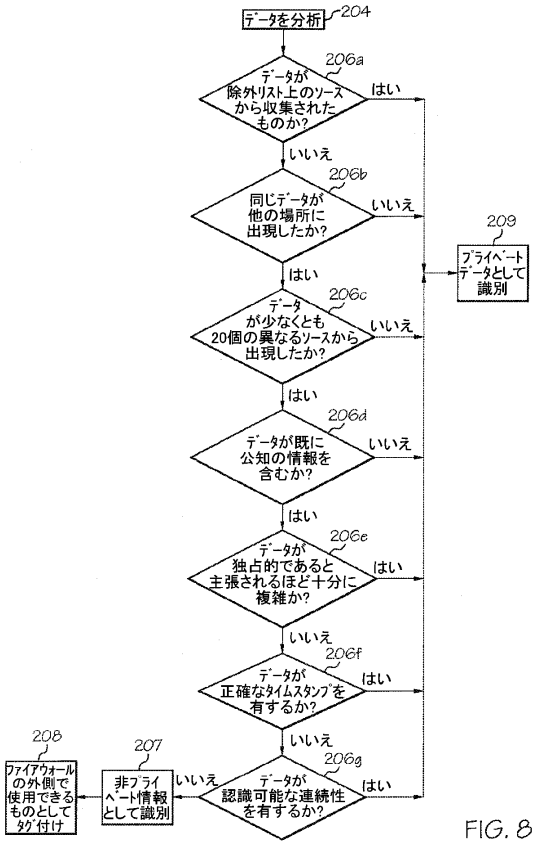

FIG. 8

フロントページの続き

(74)代理人 100086771
弁理士 西島 孝喜
(74)代理人 100109070
弁理士 須田 洋之
(74)代理人 100109335
弁理士 上杉 浩
(74)代理人 100120525
弁理士 近藤 直樹
(74)代理人 100196612
弁理士 鎌田 慎也
(72)発明者 キルガロン ウィリアム
アメリカ合衆国 オハイオ州 ４５０３６ レバノン レイク ヘヴン コート ４６２

審査官 吉田 誠

(56)参考文献 特開２０１３－１６１１０３（ＪＰ，Ａ）
特開２００１－３２５２４９（ＪＰ，Ａ）
特開２０１０－０７９４４４（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ １６／００－１６／９５８
Ｇ０６Ｆ ２１／６２