(12) STANDARD PATENT (11) Application No. **AU 2015353713 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Systems and methods for implementing a privacy firewall

(51) International Patent Classification(s)
H04L 29/06 (2006.01)

(21) Application No: **2015353713** (22) Date of Filing: **2015.11.24**

(87) WIPO No: **WO16/085897**

(30) Priority Data

(31) Number	(32) Date	(33) Country
62/084,656	**2014.11.26**	**US**

(43) Publication Date: **2016.06.02**
(44) Accepted Journal Date: **2019.10.24**

(71) Applicant(s)
LexisNexis, A Division of Reed Elsevier Inc.

(72) Inventor(s)
Kilgallon, William

(74) Agent / Attorney
Watermark Intellectual Property Pty Ltd, Level 1 109 Burwood Road, Hawthorn, VIC, 3122, AU

(56) Related Art
US 20140047551 A1
US 20050251865 A1
US 20030097594 A1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 June 2016 (02.06.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/085897 A1

(51) International Patent Classification:
H04L 29/06 (2006.01)

(21) International Application Number:
PCT/US2015/062260

(22) International Filing Date:
24 November 2015 (24.11.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/084,656 26 November 2014 (26.11.2014) US

(71) Applicant: **LEXISNEXIS, A DIVISION OF REED ELSEVIER INC.** [US/US]; 9443 Springboro Pike, Miamisburg, OH 45342 (US).

(72) Inventor: **KILGALLON, William**; 462 Lake Haven Court, Lebanon, OH 45036 (US).

(74) Agents: **PISANO, Anthony, L.** et al.; Dinsmore & Shohi LLP, 255 East Fifth Street, Suite 1900, Cincinnati, OH 45202 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: SYSTEMS AND METHODS FOR IMPLEMENTING A PRIVACY FIREWALL

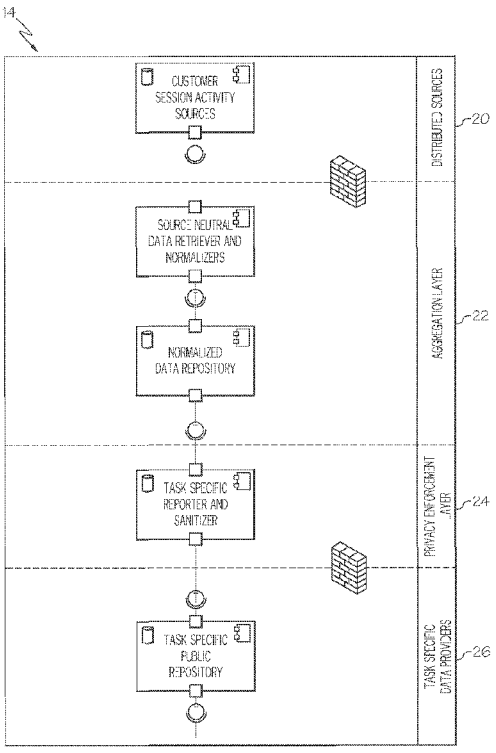

FIG. 3

(57) Abstract: Systems and methods for protecting private data behind a privacy firewall are disclosed. A system for implementing a privacy firewall to determine and provide non-private information from private electronic data includes a data storage repository, a processing device, and a non-transitory, processor-readable storage medium. The storage medium includes programming instructions that, when executed, cause the processing device to analyze a corpus of private electronic data to identify a first one or more portions of the data having non-private information and a second one or more portions of the data having private information, tag the first one or more portions of the data as allowed for use, determine whether the second one or more portions of the data includes non-private elements, and if the second one or more portions of the data comprises non-private elements, extract the non-private elements and tag the non-private elements as information allowed for use.

WO 2016/085897 A1

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

2015353713 09 Aug 2018

SYSTEMS AND METHODS FOR IMPLEMENTING A PRIVACY FIREWALL

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application Serial No. 62/084,656, entitled "SYSTEMS AND METHODS FOR DATA PRIVACY FIREWALL," filed on November 26, 2014, the entire disclosure of which is hereby incorporated by reference herein.

BACKGROUND ART

Field

The present specification generally relates to providing a privacy firewall to secure private and sensitive data and, more particularly, to systems and methods for providing access to non-private information obtained from private and sensitive data.

Technical Background

Currently, access to private and/or sensitive data may be provided by anonymizing user activity related to the data and exposing the anonymized data for viewing and use. However, such methods may fail to effectively protect privacy, as the data can be manipulated to extract private information therefrom. Such methods may also fail because the data is anonymized to the extent that it is no longer useful.

Accordingly, a need exists for systems and methods that do not anonymize private data, but rather discover and extract non-private elements therefrom in such a way that the privacy of the data is not breached, but the information contained within the private data is useful.

SUMMARY

With the foregoing in mind, one aspect of the present invention provides a system for implementing a privacy firewall to determine and provide non-private information from private electronic data to a remote computer located outside the privacy firewall, the system comprising:

2015353713 09 Sep 2019

a data storage repository comprising a corpus of the private electronic data;

a processing device; and

a non-transitory, processor-readable storage medium, wherein the non-transitory, processor-readable storage medium comprises one or more programming instructions that, when executed, cause the processing device to:

receive a query from the remote computer located outside the privacy firewall, the query comprising a request to access one or more portions of the corpus of private electronic data;

analyze the corpus of private electronic data within the privacy firewall to identify a first one or more portions of the data comprising non-private information and a second one or more portions of the data comprising private information;

tag the first one or more portions of the data as allowed for use outside the privacy firewall;

determine that the second one or more portions of the data comprises non-private elements, wherein the non-private elements are portions of the private information or additional data that is derived from the private information that are not private;

extract the non-private elements from the second one or more portions of the data, wherein a remainder of the private elements are unextracted elements;

tag the non-private elements as information allowed for use outside the privacy firewall;

tag the unextracted elements from the second one or more portions of the data as denied for use outside the privacy firewall; and

provide one or more of the first one or more portions of the data comprising non-private information and the non-private elements to the remote computer located outside of the privacy firewall.

Another aspect of the invention provides a method for implementing a privacy firewall to determine and provide non-private information from private electronic data to a remote computer located outside the privacy firewall, the method comprising:

receiving a query from the remote computer located outside the privacy firewall, the query comprising a request to access one or more portions of the corpus of private electronic data;

analyzing, by a processing device, a corpus of private electronic data contained in a storage repository within the privacy firewall to identify a first one or more portions of

2015353713 09 Sep 2019

the data comprising non-private information and a second one or more portions of the data comprising private information;

tagging, by the processing device, the first one or more portions of the data as allowed for use outside the privacy firewall;

determining, by the processing device, whether the second one or more portions of the data comprises non-private elements, wherein the non-private elements are portions of the private information or additional data that is derived from the private information that are not private;

extracting, by the processing device, the non-private elements from the second one or more portions of the data, wherein a remainder of the private elements are unextracted elements;

tagging, by the processing device, the non-private elements as information allowed for use outside the privacy firewall;

tagging, by the processing device, the unextracted elements from the second one or more portions of the data as denied for use outside the privacy firewall; and

providing, by the processing device, one or more of the first one or more portions of the data comprising non-private information and the non-private elements to the remote computer located outside of the privacy firewall.

A further aspect of the invention provides a method for implementing a privacy firewall to determine and provide non-private information from private electronic data to a remote computer located outside the privacy firewall, the method comprising:

analyzing, by a processing device, a corpus of private electronic data contained in a storage repository to identify a first one or more portions of the data comprising non-private information and a second one or more portions of the data comprising private information;

tagging, by the processing device, the first one or more portions of the data as allowed for use outside the privacy firewall;

determining, by the processing device, whether the second one or more portions of the data comprises non-private elements;

if the second one or more portions of the data comprises non-private elements:

extracting, by the processing device, the non-private elements, and

tagging, by the processing device, the non-private elements as information allowed for use outside the privacy firewall; and

2015353713 09 Sep 2019

providing, by the processing device, one or more of the first one or more portions of the data comprising non-private information and the non-private elements to the remote computer,

wherein determining whether the second one or more portions of the data comprises non-private elements comprises:

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that have appeared in other locations;

if the one or more sub-portions have appeared in other locations, tagging, by the processing device, the elements as non-private elements;

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that have appeared from a threshold number of distinct sources;

if the one or more elements have appeared from at least the threshold number of distinct sources, tagging, by the processing device, the one or more elements as non-private elements;

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that contain information that is already public knowledge;

if the one or more elements contain information that is already public knowledge, tagging, by the processing device, the one or more sub-portions as non-private elements;

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that are complex such that the one or elements are claimable as proprietary information;

if the one or more elements are complex, tagging, by the processing device, the one or more elements as denied for use outside the privacy firewall;

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that have recognizable sequences;

if the one or more elements have recognizable sequences, tagging, by the processing device, the one or more elements as denied for use outside the privacy firewall;

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that have precise timestamps; and

if the one or more elements have precise timestamps, tagging, by the processing device, the one or more elements as denied for use outside the privacy firewall.

2015353713 09 Sep 2019

A still further aspect of the invention provides a system for providing non-private information from private electronic data through a privacy firewall to a remote computer located outside the privacy firewall, the system comprising:

a data storage repository behind the privacy firewall, the data storage repository comprising a corpus of private electronic data that has been tagged as non-private information, private information not to be distributed, non-private elements of private information to be distributed, or depersonalized private information, wherein the non-private elements are portions of the private information or additional data that is derived from the private information that are not private;

a processing device; and

a non-transitory, processor-readable storage medium, wherein the non-transitory, processor-readable storage medium comprises one or more programming instructions that, when executed, cause the processing device to:

receive a search string from the remote computer located outside the privacy firewall, wherein the search string comprises a question;

search the data storage repository for one or more portions of the private electronic data that correspond to the search string;

if the one or more portions of the electronic data contain the non-private information, non-private elements of the private information to be distributed, or the depersonalized private information, provide a response to the remote computer, wherein the response contains information contained within the one or more portions of the electronic data contain the non-private information, the non-private elements of the private information to be distributed, or the depersonalized private information; and

tagging a remainder of the private information that is not the non-private information, non-private elements of the private information to be distributed, or the depersonalized private information as denied for use.

In one embodiment, a system for implementing a privacy firewall to determine and provide non-private information from private electronic data includes a data storage repository having a corpus of private electronic data, a processing device, and a non-transitory, processor-readable storage medium. The non-transitory, processor-readable

[continued on page 2]

storage medium includes one or more programming instructions that, when executed, cause the processing device to analyze the corpus of electronic data to identify a first one or more portions of the data having non-private information and a second one or more portions of the data having private information, tag the first one or more portions of the data as allowed for use outside the privacy firewall, determine whether the second one or more portions of the data includes non-private elements, and if the second one or more portions of the data comprises non-private elements, extract the non-private elements and tag the non-private elements as information allowed for use outside the privacy firewall.

In another embodiment, a method for implementing a privacy firewall to determine and provide non-private information from private electronic data includes analyzing, by a processing device, a corpus of private electronic data contained in a storage repository to identify a first one or more portions of the data comprising non-private information and a second one or more portions of the data including private information, tagging, by the processing device, the first one or more portions of the data as allowed for use outside the privacy firewall, determining, by the processing device, whether the second one or more portions of the data includes non-private elements, and if the second one or more portions of the data comprises non-private elements, extracting, by the processing device, the non-private elements and tagging, by the processing device, the non-private elements as information allowed for use outside the privacy firewall.

In yet another embodiment, a system for providing non-private information from private electronic data through a privacy firewall includes a data storage repository behind the privacy firewall, a processing device, and a non-transitory, processor-readable storage medium. The data storage repository includes a corpus of private electronic data that has been tagged as non-private information, private information not to be distributed, non-private elements of private information to be distributed, or depersonalized private information. The non-transitory, processor-readable storage medium includes one or more programming instructions that, when executed, cause the processing device to receive search string from a user outside the privacy firewall, search

the data storage repository for one or more portions of the electronic data that correspond to the search string, and if the one or more portions of the electronic data contain the non-private information, the non-private elements of the private information to be distributed, or the depersonalized private information, providing a response to the search string. The search string includes a question, and the response contains information contained within the one or more portions of the electronic data contain the non-private information, the non-private elements of the private information to be distributed, or the depersonalized private information.

These and additional features provided by the embodiments described herein will be more fully understood in view of the following detailed description, in conjunction with the drawings.

BRIEF DESCRIPTION OF DRAWINGS

The embodiments set forth in the drawings are illustrative and exemplary in nature and not intended to limit the subject matter defined by the claims. The following detailed description of the illustrative embodiments can be understood when read in conjunction with the following drawings, wherein like structure is indicated with like reference numerals and in which:

FIG. 1 depicts a schematic depiction of an illustrative computing network for a system for providing access to data behind a privacy firewall according to one or more embodiments shown and described herein;

FIG. 2 depicts a schematic depiction of the server computing device from FIG. 1, further illustrating hardware and software that may be used in providing data according to one or more embodiments shown and described herein;

FIG. 3 depicts a schematic depiction of the various layers of the privacy firewall from FIG. 1 according to one or more embodiments shown and described herein;

FIG. 4 depicts a flow diagram of an illustrative method of providing data in response to a request according to one or more embodiments shown and described herein;

FIG. 5 depicts a schematic depiction of an illustrative search input screen of a graphical user interface according to one or more embodiments shown and described herein;

FIG. 6 depicts a schematic depiction of an illustrative search input screen of a graphical user interface containing auto-complete options according to one or more embodiments shown and described herein;

FIG. 7 depicts a flow diagram of an illustrative method of analyzing and categorizing data according to one or more embodiments shown and described herein; and

FIG. 8 depicts a flow diagram of an illustrative method of determining whether private data contains non-private information according to one or more embodiments shown and described herein.

DESCRIPTION OF EMBODIMENTS

Referring generally to the figures, embodiments described herein are directed to systems and methods for implementing a privacy firewall to restrict access to a corpus of private data stored on a server located within the bounds of the privacy firewall. In particular embodiments, the systems and methods described herein may generally be implemented to ensure that the private data remains secured, while still providing informational answers in response to questions submitted by one or more users, where the answers are obtained from the private data. It may generally be assumed that all data behind the privacy firewall is private data. However, certain portions of the data may actually contain non-private information. In addition, the remaining portions of the data that contain private information may also contain non-private elements. The non-private information and the non-private elements from the private information may be tagged as usable to answer the question submitted by the user while preserving the privacy of the data.

The methods and systems disclosed herein may be used, for example, in instances where an uncertainty may exist as to whether the use of data will "cross users," (i.e.,

whether an entity's private data will be seen or deduced by anyone other than the entity) or in cases where it is desired to provide a data repository that does not need highly controlled access. Non-limiting examples of instances where the privacy firewall may be applied include data derived from person A that will only ever effect person A and the data repository access is tightly controlled, data derived from person A that can affect person B or the data repository access is not tightly controlled, data that can be mapped to a user that performed a particular action (such as via user ID or internet protocol (IP) address), data that must adhere to a search string retention policy (e.g., data that must be removed or depersonalized within a particular time period), data that must adhere to privacy standards, and data that is lightly depersonalized (i.e., the user ID or IP address has been removed).

As used herein, the term "non-private information" refers to data that contains information to which an individual or entity would not have any expectation of privacy. The non-private information may be stored in repositories where all data stored therein is initially assumed to be private. As such, a determination may be made that the data contains non-private information. An illustrative example of non-private information as used herein is data that is very commonly used across widely disparate sources. Another illustrative example of non-private information is data that is not specifically related to a particular individual or entity. Yet another illustrative example of non-private information is data relating to search narrowing constructs from a search string, such as demographics, data source, historical intervals, geographic scope, and/or the like. Yet another illustrative example of non-private information is data relating to non-private web browsing activity, such as published information that is readily available to anyone on the Internet, publicly available electronic folders containing data, and/or the like. In some embodiments, non-private information may also be referred to as non-sensitive data. In some embodiments, a determination of whether certain data is deemed non-private may be completed by applying one or more rules.

As used herein, the term "private data" refers to data that contains information to which an individual or an entity would have an expectation of privacy. Illustrative

examples of private data may include, but are not limited to, data relating to personal information for a particular individual or entity, data that can be mapped to a particular individual or entity (such as data that contains a particular user ID, IP address, or the like), data that is subject to a retention policy, data that is considered private because of a particular privacy standard, regulatory requirement, or the like (such as data deemed private by the Health Insurance Portability and Accountability Act (HIPAA) and/or the like), data that can only be derived from a particular individual, entity, or a specific groups of individuals and/or entities, complex data that could be claimed as being proprietary, data that contains information not generally known to the public, and data that would allow someone to reconstruct the information contained therein to gain further intelligence that could compromise a particular individual's or entity's privacy. In some embodiments, a determination of whether certain data is deemed private may be completed by applying one or more rules. In some embodiments, all data stored in a repository behind a privacy firewall may initially be presumed to be private data until it is analyzed to determine whether non-private information is included.

Even though the data in the repositories described herein may initially be deemed private, certain portions of the data or additional data that is derived from the private data may be categorized herein as "non-private elements" of private data. Illustrative examples of non-private elements may include, but are not limited to, data that has been depersonalized (such as data that has had personal information removed therefrom), identical data has appeared from a particular number of unique locations (such as searches that were conducted from a particular number of unique IP addresses), identical data that has appeared from a particular number of unique individuals and/or entities, non-private information stored in a private repository, and non-identifying metadata obtained from private data. The non-identifying metadata may include, but is not limited to, a geographic region of the individual or entity that generated the private data (such as a state, province, region, or territory, but no more specific than such), a time stamp indicating a date and an hour the data was generated (but not a minute or a second), certain search terms and connectors related to the private data, a market segment related to the private data, a product used to search (such as a particular web browser, search

engine, or the like), and a number of hits from search results as a result of a search. The rules for identifying non-private information residing in a private repository may not completely protect user privacy if the data published is too specific. For example, if a common search string is internally published, but it also includes the source IP address of a user that completed the search, the privacy of the user may be compromised.

A spectrum of privacy may exist regarding user data. For example, at one extreme, a complete search string that includes the name of the individual conducting the search may be deemed private and/or sensitive data. At the other extreme, just because a user somewhere used the letter "e" in a search does not mean that another can never allow the letter "e" to appear in anything used and published. Between the two extremes, a middle ground may exist where data ceases to have any kind of privacy implications. For example, if 10,000 different users have entered the search term "Roe v. Wade," and then continued on to browse U.S. Supreme Court case citation 410 U.S. 113, an invasion of privacy likely does not exist if a user interface (UI) feature is provided that offers to jump directly to 410 U.S. 113 when a user begins typing in "Roe v.", even though the intelligence to do this might have been related to past actions of specific users.

Referring now to the drawings, FIG. 1 depicts an illustrative computing network that depicts components for a system for providing a privacy firewall that determines non-private information in a private data repository, provides responses to questions based on the non-private information, and/or auto-completes search requests based on the non-private information, according to embodiments shown and described herein. As illustrated in FIG. 1, a computer network 10 may include a wide area network (WAN), such as the Internet, a local area network (LAN), a mobile communications network, a public service telephone network (PSTN), a personal area network (PAN), a metropolitan area network (MAN), a virtual private network (VPN), and/or another network. The computer network 10 may generally be configured to electronically connect one or more computing devices and/or components thereof. Illustrative computing devices may include, but are not limited to, a user computing device 12a, a server computing device 12b, and an administrator computing device 12c.

The user computing device 12a may generally be used as an interface between the user and the other components connected to the computer network 10. Thus, the user computing device 12a may be used to perform one or more user-facing functions, such as receiving one or more inputs from a user or providing information to the user, as described in greater detail herein. Additionally, included in FIG. 1 is the administrator computing device 12c. In the event that the server computing device 12b requires oversight, updating, or correction, the administrator computing device 12c may be configured to provide the desired oversight, updating, and/or correction. The administrator computing device 12c may also be used to input additional data into the corpus stored on the server computer device 12b.

The server computing device 12b may receive data from one or more sources, store the data, and provide information from certain portions of the data to the user computing device 12a in the form of answers to questions or auto-complete suggestions when access to such information is authorized and the information is tagged as allowed for distribution. A determination of whether the information is allowed for distribution may generally be completed by a privacy firewall 14 that sits between the server computing device 12b and the computer network 10. Thus, the privacy firewall 14 (which may also be referred to as a privacy veil) may allow or deny access to certain information from the data stored at the server computing device 12b, as described in greater detail herein.

It should be understood that while the user computing device 12a and the administrator computing device 12c are depicted as personal computers and the server computing device 12b is depicted as a server, these are nonlimiting examples. More specifically, in some embodiments, any type of computing device (e.g., mobile computing device, personal computer, server, etc.) may be used for any of these components. Additionally, while each of these computing devices is illustrated in FIG. 1 as a single piece of hardware, this is also merely an example. More specifically, each of the user computing device 12a, server computing device 12b, and administrator

computing device 12c may represent a plurality of computers, servers, databases, components, and/or the like.

FIG. 2 depicts the server computing device 12b, from FIG. 1, further illustrating a system for determining non-private information, searching a document corpus, generating responses to questions presented by users, and/or generating auto-complete suggestions. In addition, the server computing device 12b may include a non-transitory computer-readable medium for searching a document corpus or generating search queries embodied as hardware, software, and/or firmware, according to embodiments shown and described herein. While in some embodiments the server computing device 12b may be configured as a general purpose computer with the requisite hardware, software, and/or firmware, in some embodiments, the server computing device 12b may also be configured as a special purpose computer designed specifically for performing the functionality described herein.

As also illustrated in FIG. 2, the server computing device 12b may include a processor 30, input/output hardware 32, network interface hardware 34, a data storage component 36 (which may store non-private information 38a, non-private elements of private data 38b, and other data 38c), and a non-transitory memory component 40. The memory component 40 may be configured as volatile and/or nonvolatile computer readable medium and, as such, may include random access memory (including SRAM, DRAM, and/or other types of random access memory), flash memory, registers, compact discs (CD), digital versatile discs (DVD), and/or other types of storage components. Additionally, the memory component 40 may be configured to store operating logic 42 and search logic 44 (each of which may be embodied as a computer program, firmware, or hardware, as an example). A local interface 46 is also included in FIG. 2 and may be implemented as a bus or other interface to facilitate communication among the components of the server computing device 12b.

The processor 30 may include any processing component configured to receive and execute instructions (such as from the data storage component 36 and/or memory component 40). The input/output hardware 32 may include a monitor, keyboard, mouse,

printer, camera, microphone, speaker, touch-screen, and/or other device for receiving, sending, and/or presenting data. The network interface hardware 34 may include any wired or wireless networking hardware, such as a modem, LAN port, wireless fidelity (Wi-Fi) card, WiMax card, mobile communications hardware, and/or other hardware for communicating with other networks and/or devices.

It should be understood that the data storage component 36 may reside local to and/or remote from the server computing device 12b and may be configured to store one or more pieces of data and selectively provide access to the one or more pieces of data. As illustrated in FIG. 2, the data storage component 36 may store non-private information 38a, non-private elements of private data 38b, and other data 38c, as described in greater detail herein.

Included in the memory component 40 are the operating logic 42 and the search logic 44. The operating logic 42 may include an operating system and/or other software for managing components of the server computing device 12b. The search logic 44 may be configured to generate search queries from user-input within the graphical user interface, as described in detail below.

It should be understood that the components illustrated in FIG. 2 are merely illustrative and are not intended to limit the scope of this disclosure. More specifically, while the components in FIG. 2 are illustrated as residing within the server computing device 12b, this is a nonlimiting example. In some embodiments, one or more of the components may reside external to the server computing device 12b. Similarly, while FIG. 2 is directed to the server computing device 12b, other components such as the user computing device 12a and the administrator computing device 12c may include similar hardware, software, and/or firmware.

FIG. 3 depicts various layers of the privacy firewall 14 from FIG. 1. The layers depicted in FIG. 3 are merely illustrative. Thus, fewer or additional layers may be used without departing from the scope of the present disclosure. In addition, certain layers may be collapsed or further stratified into additional layers. Each layer may represent an amount of access to the data contained within the server computing device 12b (FIG. 1)

that is provided to an outside requester, such as, for example, a user of the user computing device 12a (FIG. 1). The access may generally be in the form of an answer to a question posed by a user or as an auto-complete suggestion, rather than direct access to the data. Illustrative layers may include, for example, a distributed sources layer 20, an aggregation layer 22, a privacy enforcement layer 24, and a task specific data provider layer 26. In some embodiments, the distributed sources layer 20, the aggregation layer 22, and the privacy enforcement layer 24 may be highly restrictive data layers where little or no data categorized by such layers is allowed to be accessed without tuning the data, as described in greater detail herein. In some embodiments, the task specific data provider layer 26 may be a low restrictive data layer where more or all of the data categorized by the layer is allowed to be accessed.

The distributed sources layer 20 may categorize data that is typically found in, for example, customer session activity sources. Such customer session activity sources may represent a plurality of sources present on a plurality of different platforms and/or applications from which data is received and stored. As a nonlimiting example, one source may be a search box of a particular program or application. Data from the different sources may be stored in different data repositories. Certain data repositories may have more data restrictions than other data repositories. As such, the data may not be normalized across the different repositories.

The aggregation layer 22 may categorize data that is typically found in, for example, normalized data repositories. That is, the data may have been obtained from various data repositories, each in a respective native format, and normalized by one or more normalization tools ("normalizers") into a single consistent format. In some embodiments, as the data is normalized, it may be anonymized as much as practical so that the amount of sensitive information contained in the data is minimized. However, in some embodiments, the repository may contain a common and consistent set of attributes that are sufficient for certain modules and a full anonymization may not be possible until a particular query is presented (i.e., a query that relates to the data). Thus, despite the

anonymization, the data contained in the repository may still be highly sensitive. Accordingly, access to such data may be highly restricted.

The privacy enforcement layer 24 may categorize data that has been passed through a task specific reporter and sanitizer module. Such data may be, for example, normalized data which can contain sensitive information if a query is presented a certain way. For example, a privacy veil may be pierced by asking a plurality of questions in a plurality of different ways and using Boolean algebra operations to tease out more data than was intended to be exposed. Thus, it may be desirable to strictly limit the types of questions that can be asked to obtain certain information and also ensure the total variety of questions that are exposed cannot have their answers combined to leak information. Therefore, each query that is presented to the privacy firewall 14 may have a specific module created to deliver results. Each specific module may be vetted and built in a highly restricted environment such that the module acts as a bridge to translating sensitive data to non-sensitive data.

The task specific data provider layer 26 may generally include a task specific public repository of information that is generally non-private, desensitized private data, or non-private elements of private data. Such data may be used to answer a query.

Referring now to FIG. 4, a flow diagram that illustrates a method of implementing a privacy firewall according to one or more embodiments is provided. As described in greater detail herein, embodiments may enable a user to request information and view non-private information and/or non-private elements of private data. At block 180, the system may generate a graphical user interface for display on a display device of the user computing device 12a. The graphical user interface is configured such that a user can send a search string at block 182. Referring to FIG. 5, the graphical user interface may comprise a search string input screen 100 that is configured to receive a search string from a user. It should be understood that embodiments are not limited to the configurations of the graphical user interfaces illustrated throughout the figures, and that other graphical user interface configurations are possible. In one embodiment, the

network 10 is the Internet and the graphical user interfaces described herein are presented to the user via a web browser.

The search string input screen 100 comprises a search string field 103 into which a user may enter one or more terms that make up a desired search string (e.g., by the use of a keyboard). In one embodiment, the search string may be a natural language search string. For example, the user may ask a question, such as, for example, "what is the landmark case from the Supreme Court regarding abortions?" In another example, such as the embodiment illustrated in FIG. 5, the search string "Roe v. Wade" has been entered into the search string field 103, as the particular user may be interested in searching for information relating to the U.S. Supreme Court decision from 1973. As shown in FIG. 6, in some embodiments, the user may only need to enter one or more letters in the search string field 103, and the system may generate suggested auto-complete options based on data obtained from the private data repositories contained within the server computing device 12b. For example, if the user is searching a case law database, when the user types the letters "RO," certain auto complete options that contain the letters "RO" may be presented to the user, such as, for example, Roe v. Wade, In Re Ross, and/or the like, as shown below the user's input in the search string field 104. The auto-complete suggestions may be generated, for example, as a result of data indicating that many users have searched for those particular cases, which may optionally be ranked by popularity. Thus, the system may attempt to guess what a user might want to search for before the user has completed typing the search string based on the private data contained within the server computing device 12b. However, the information that is obtained to provide the auto-complete suggestions may generally be the non-private information and/or non-private elements of the private data such that terms that may breach one's privacy are not included in the auto complete options.

As shown in FIGS. 5 and 6, the search string input screen 100 may also include other input features, such as an option button 106, a section filter input 102, and a search initiation icon 105. It should be understood that more or fewer input features may be used. In the example illustrated in FIGS. 5 and 6, the option button 106 allows a user to

also search machine translations of electronic data that is not in the native language of the corpus in which the user is searching. Other options may also be provided. The section filter input 102 may enable a user to only search particular sections or combinations of sections of the data. For example, in the case law context, the user may use the section filter input 102 to search only the case summary section, a keywords section, a judicial opinion section, a facts section, and/or the like.

The user may initiate a search based on the search string entered into the search string field 103 by clicking or otherwise selecting the search initiation icon 105. Individual terms of the search string are used as query terms to analyze the search string at block 184. Analyzing the search string may generally include determining what the user is searching for, which may be completed by any method now known or later developed. In some embodiments, an appropriate search query may be determined in block 186 based on the user's submitted search string. That is, the search string submitted by the user may be received and interpreted, and an appropriate search query may be generated based on the data to be accessed, the type of system containing the data, and/or the like. The appropriate search query may be generated using any number of query generation techniques. For example, a Boolean weighted search query may be generated based on the analysis of the search string submitted by the user.

In block 188, the data contained in a repository behind the firewall 14, such as in the server computing device 12b, may be searched. If data relating to the search query is discovered, the data may be analyzed to determine whether the data is private or non-private in step 190. In some embodiments, the private data may be further analyzed to determine whether it contains non-private elements in step 192. In addition, the private data may be analyzed to determine whether certain elements may be tuned to make those elements non-private elements in step 194. If the data can be tuned, such tuning may be completed in step 196 to obtained tuned data (which may also be referred to as depersonalized data). For example, the data may be tuned to adjust the granularity of a time and date stamp, adjust the granularity of a geographical location, and/or the like. In some embodiments, the granularity may be adjusted above a particular granularity

threshold. A geographic granularity threshold may be, for example, between a state or province and a city (e.g., "above" the threshold may be granularity of a state, a province, a country, a region, or the like, and "below" the threshold may be additional detail such as city, street address, or the like). A time granularity threshold may be, for example, between an hour and a minute (e.g., time displayed in hours, days, weeks, months, and years may be "above" the threshold, whereas time displayed in minutes and seconds may be "below" the threshold). In some embodiments, steps 190, 192, 194, and 196 may be completed for each search query. In other embodiments, steps 190, 192, 194, and 196 may be completed as soon as data is obtained in the repository so as to allow the searching in step 188 to be quickly completed.

The query terms of the search query are used to search the corpus of data to provide a response to the search string in step 198. The response is generally derived from the non-private information and/or the non-private elements of the private data, including the tuned portions of the private data (if any). The response may be a natural language answer that responds to the question asked by the user, one or more links to particular references, a set of returned electronic documents for display, and/or the like.

Referring again to FIG. 1, in various embodiments, the server computing device 12b may continue to gather new data as it is generated and/or made available. The new data may be analyzed such that a determination can be made as to whether the data contains private or sensitive data such that private or sensitive information is not allowed past the firewall 14, as described herein. Thus, a process for analyzing and categorizing the data is depicted in FIG. 7.

As shown in FIG. 7, and also referring to FIG. 1, the data may be obtained from a repository in step 202. For example, in some embodiments, the data may be copied or moved from a remote repository to the server computing device 12b behind the firewall 14. In other embodiments, the repository may be located within the server computing device 12b (e.g., in the data storage 36 (FIG. 2)) or otherwise behind the firewall 14, thereby making a copy or a transfer unnecessary.

As described herein, the data may initially be presumed to be private data. In step 204, the data may be analyzed to determine in step 206 whether the data contains non-private information. If the data does not contain any portion that could be private, the data may be identified as non-private information and may be tagged as allowed for use outside the privacy firewall 14 in step 208. For example, if the data entirely contains publicly available information such as unsealed court records, publicly-available real estate records, and/or the like, the data may be determined in step 206 to not contain private data and tagged as allowed for use outside the privacy firewall 14 in step 208.

On the other hand, if any portion of the data contains information that is sensitive, private, or could be viewed as sensitive or private, the data may continue to be identified as private data. For example, if the data contains court records, but the court records contain private information such as an individual's home address, the data may be identified as private data.

As shown in FIG. 8, the determination of whether the data contains non-private information may include a plurality of determination steps. Such steps are merely illustrative, and it should be understood that alternative, additional, or fewer steps may be completed without departing from the scope of the present disclosure. Further, it should be understood that embodiments are not limited by this order of steps as shown in FIG. 8. One illustrative step, as shown in step 206a, a determination may be made as to whether the data was gathered from a source that is on an excluded list. Illustrative excluded lists may include, for example, lists containing information designated private per HIPAA, the Healthcare Interoperability Testing and Conformance Harmonisation (HITCH) project, the Driver's Privacy Protection Act (DPPA), the Gramm-Leach-Bliley Act (GLBA) (also known as the Financial Services Modernization Act of 1999), the Payment Card Industry Data Security Standard (PCI DSS), and/or the like. If the data is gathered from sources on an excluded list, the data may be identified as private in step 209. If the data is not gathered from sources on an excluded list, additional determination steps may be completed or the data may be identified as containing non-

private information in step 207 and tagged as being allowed for use outside the privacy firewall in step 208.

In illustrative step 206b, a determination may be made as to whether identical data has appeared in other locations. For example, if data is gathered regarding a particular search string that was entered at a search engine interface, the determination in step 206b may look to see if that same search string was entered from another location. Locations may be determined by reviewing metadata or the like. For example, the metadata regarding a particular search query may include the IP address of the device used to enter the search string in the search engine interface. If an identical search query was entered by different IP addresses, and such IP addresses constitute different locations (i.e., not IP addresses originating from the same physical location), it may be determined that the data has appeared in other locations. If the data has not appeared in other locations, the data may be identified as private in step 209. If the data has appeared in other locations, additional determination steps may be completed or the data may be identified as containing non-private information in step 207 and tagged as being allowed for use outside the privacy firewall in step 208.

In illustrative step 206c, a determination may be made as to whether the data has been obtained from at least 20 distinct sources. For example, if the data contains metadata indicative of a source of the data (e.g., an IP address or the like), the determination may include reviewing the metadata to ensure that the data was obtained from 20 distinct sources. It should be understood that the number of sources used herein (20) is merely illustrative, and that any number of sources may be specified, particularly a number of sources that ensures that the date is not private. For example, the number of sources may vary based on the type of data, certain rules or policies regarding the data, and/or the like. If the data has not appeared from at least 20 distinct sources, the data may be identified as private in step 209. If the data has appeared from at least 20 distinct sources, additional determination steps may be completed or the data may be identified as containing non-private information in step 207 and tagged as being allowed for use outside the privacy firewall in step 208.

In illustrative step 206d, a determination may be made as to whether the data contains information that is already public knowledge. For example, if the data contains information that would ordinarily be private, such as an individual's home address, but the individual has publicly broadcasted his home address to advertise a business that he operates out of his home, such information may be considered to be already public knowledge. If the data contains information that is not already public knowledge, the data may be identified as private in step 209. If the data contains information that is already public knowledge, additional determination steps may be completed or the data may be identified as containing non-private information in step 207 and tagged as being allowed for use outside the privacy firewall in step 208.

In illustrative step 206e, a determination may be made as to whether the data is complex enough to be claimed as proprietary. The complexity of the data may be made on a case-by-case basis that is based on the nature of the data, the nature of the sources of the data, the context within which the data was gathered, and any agreements made with a provider of the data (e.g., a terms of use agreement). For example, if an entity develops a complex algorithm that could be claimed as proprietary (such as, for example, a trade secret) and the algorithm appears in the data, at least the portion of the data containing the algorithm may be considered private. In another example, if an entity enters a search string that is very narrowly construed and/or unique, the search string may be considered complex. Accordingly, if the data is determined to be complex enough to be claimed as proprietary, the data may be identified as private data in step 209. If the data is not complex enough to be claimed as proprietary, additional determination steps may be completed or the data may be identified as containing non-private information in step 207 and tagged as being allowed for use outside the privacy firewall in step 208.

In illustrative step 206f, a determination may be made as to whether the data contains precise timestamps. For example, if the data relates to search strings having precise timestamps that may allow one to identify the individual or entity that conducted the search, the data may be private. Unless the data is appropriately tuned to adjust the

granularity of the timestamp (e.g., no more specific than the hour the search string was inputted) as described herein, the data may be identified as private in step 209. If the data does not contain precise timestamps, or if it has been appropriately tuned, additional determination steps may be completed or the data may be identified as containing non-private information in step 207 and tagged as being allowed for use outside the privacy firewall in step 208.

In illustrative step 206g, a determination may be made as to whether the data contains recognizable sequences. A recognizable sequence may generally be a sequence of data strings that, when observed together, contain information that should be private, even if the data strings standing alone would not be considered private. For example, if the data contains a sequence that might allow one to determine the individual entity or data associated with the data, such data may be identified as private data in step 209. In another example, if an entity conducts 4 subsequent searches using the terms "favorable courts for manufacturing-based class action lawsuits," "companies making scraper widgets," "companies in the state of Connecticut," and "individuals admitted to emergency rooms in Connecticut for cuts," it may be deduced that the entity is an attorney or law firm considering filing a class-action lawsuit against scraper widget makers in Connecticut, which may constitute private information even though the search strings, taken individually, would not reveal such information and may not be private. If the data does not contain any sequences that may be used to identify the individual or entity associated with the data, additional determination steps may be completed or the data may be identified as containing non-private information in step 207 and tagged as being allowed for use outside the privacy firewall in step 208.

Referring again to FIG. 7, in step 210, a determination may be made as to whether the private data contains non-private elements. That is, if the private data has certain portions that, when standing alone, would constitute non-private information, such private data is identified as containing private portions and the non-private elements may be extracted therefrom in step 214. As previously described herein, the non-private elements may be extracted in such a way that private information may not be gleaned

from the non-private elements. For example, if the non-private element is metadata that can be extracted from the private data, such metadata must be restricted in scope (e.g., a geographic location of a user may not be any more specific than the state, territory, province, etc. in which the user is located). The non-private elements upon being extracted, may be tagged as being allowed for use outside the privacy firewall 14 in step 216. If the data does not contain non-private elements, the data may be tagged in step 212 as being denied data to which access is not provided through the privacy firewall 14.

It should be understood that embodiments described herein provide for systems and methods for providing a privacy firewall that allows access to private data without breaching the privacy of an individual or entity that generated the data. The nature of the systems and methods herein allow a user to ask a question or type a search string that may require private data access to obtain an answer. The user may then receive a response to the question or may be provided with auto-complete suggestions without ever obtaining access to private data repositories.

While particular embodiments have been illustrated and described herein, it should be understood that various other changes and modifications may be made without departing from the spirit and scope of the claimed subject matter. Moreover, although various aspects of the claimed subject matter have been described herein, such aspects need not be utilized in combination. It is therefore intended that the appended claims cover all such changes and modifications that are within the scope of the claimed subject matter.

2015353713 09 Sep 2019

CLAIMS:

1. A system for implementing a privacy firewall to determine and provide non-private information from private electronic data to a remote computer located outside the privacy firewall, the system comprising:

a data storage repository comprising a corpus of the private electronic data;

a processing device; and

a non-transitory, processor-readable storage medium, wherein the non-transitory, processor-readable storage medium comprises one or more programming instructions that, when executed, cause the processing device to:

receive a query from the remote computer located outside the privacy firewall, the query comprising a request to access one or more portions of the corpus of private electronic data;

analyze the corpus of private electronic data within the privacy firewall to identify a first one or more portions of the data comprising non-private information and a second one or more portions of the data comprising private information;

tag the first one or more portions of the data as allowed for use outside the privacy firewall;

determine that the second one or more portions of the data comprises non-private elements, wherein the non-private elements are portions of the private information or additional data that is derived from the private information that are not private;

extract the non-private elements from the second one or more portions of the data, wherein a remainder of the private elements are unextracted elements;

tag the non-private elements as information allowed for use outside the privacy firewall;

tag the unextracted elements from the second one or more portions of the data as denied for use outside the privacy firewall; and

provide one or more of the first one or more portions of the data comprising non-private information and the non-private elements to the remote computer located outside of the privacy firewall.

2015353713 09 Sep 2019

2. The system of claim 1, wherein the one or more programming instructions that, when executed, cause the processing device to analyze the corpus of private electronic data further cause the processing device to:

determine whether the corpus of private electronic data was obtained from sources on an excluded list; and

if the corpus of private electronic data was obtained from sources on the excluded list, tag the corpus of private electronic data as denied for use outside the privacy firewall.

3. The system of claim 1, wherein the one or more programming instructions that, when executed, cause the processing device to determine whether the second one or more portions of the data further cause the processing device to:

determine whether the second one or more portions of the data comprises one or more elements that have appeared in other locations; and

if the one or more sub-portions have appeared in other locations, tag the elements as non-private elements.

4. The system of claim 1, wherein the one or more programming instructions that, when executed, cause the processing device to determine whether the second one or more portions of the data comprises non-private elements further cause the processing device to:

determine whether the second one or more portions of the data comprises one or more elements that have appeared from a threshold number of distinct sources; and

if the one or more elements have appeared from at least the threshold number of distinct sources, tag the one or more elements as non-private elements.

5. The system of claim 1, wherein the one or more programming instructions that, when executed, cause the processing device to determine whether the second one or more portions of the data comprises non-private elements further cause the processing device to:

determine whether the second one or more portions of the data comprises one or more elements that contain information that is already public knowledge; and

if the one or more elements contain information that is already public knowledge, tag the one or more elements as non-private elements.

2015353713 09 Sep 2019

6. The system of claim 1, wherein the one or more programming instructions that, when executed, cause the processing device to determine whether the second one or more portions of the data further cause the processing device to:

determine whether the second one or more portions of the data comprises one or more elements that are complex such that the one or elements are claimable as proprietary information; and

if the one or more elements are complex, tag the one or more elements as denied for use outside the privacy firewall.

7. The system of claim 1, wherein the one or more programming instructions that, when executed, cause the processing device to determine whether the second one or more portions of the data comprises non-private elements further cause the processing device to:

determine whether the second one or more portions of the data comprises one or more sub-portions that have recognizable sequences; and

if the one or more sub-portions have recognizable sequences, tag the one or more sub-portions as denied for distribution outside the privacy firewall.

8. The system of claim 1, wherein the one or more programming instructions that, when executed, cause the processing device to determine whether the second one or more portions of the data comprises non-private elements further cause the processing device to:

determine whether the second one or more portions of the data comprises one or more elements that have precise timestamps; and

if the one or more elements have precise timestamps, tag the one or more elements as denied for use outside the privacy firewall.

9. The system of claim 1, wherein the one or more programming instructions that, when executed, cause the processing device to determine whether the second one or more portions of the data comprises non-private elements further cause the processing device to:

determine whether the second one or more portions of the data comprises one or more elements that have a timestamp that is below a granularity threshold; and

2015353713 09 Sep 2019

if the timestamp is below the granularity threshold:

tune the one or more elements to a time range above the granularity threshold, and

tag the one or more tuned elements as non-private elements.

10. The system of claim 1, wherein the one or more programming instructions that, when executed, cause the processing device to determine whether the second one or more portions of the data comprises non-private elements further cause the processing device to:

determine whether the second one or more portions of the data comprises one or more elements that have a geographic location that is below a granularity threshold; and

if the geographic location is below the granularity threshold:

tune the one or more elements to a geographic location above the granularity threshold, and

tag the one or more tuned elements as non-private elements.

11. The system of claim 1, further comprising one or more programming instructions that, when executed, cause the processing device to:

provide a response to the query to a user via a user interface, wherein the response to the query contains information obtained from one or more of the first one or more portions of the data comprising non-private information and the non-private elements.

12. A method for implementing a privacy firewall to determine and provide non-private information from private electronic data to a remote computer located outside the privacy firewall, the method comprising:

receiving a query from the remote computer located outside the privacy firewall, the query comprising a request to access one or more portions of the corpus of private electronic data;

analyzing, by a processing device, a corpus of private electronic data contained in a storage repository within the privacy firewall to identify a first one or more portions of the data comprising non-private information and a second one or more portions of the data comprising private information;

2015353713 09 Sep 2019

tagging, by the processing device, the first one or more portions of the data as allowed for use outside the privacy firewall;

determining, by the processing device, whether the second one or more portions of the data comprises non-private elements, wherein the non-private elements are portions of the private information or additional data that is derived from the private information that are not private;

extracting, by the processing device, the non-private elements from the second one or more portions of the data, wherein a remainder of the private elements are unextracted elements;

tagging, by the processing device, the non-private elements as information allowed for use outside the privacy firewall;

tagging, by the processing device, the unextracted elements from the second one or more portions of the data as denied for use outside the privacy firewall; and

providing, by the processing device, one or more of the first one or more portions of the data comprising non-private information and the non-private elements to the remote computer located outside of the privacy firewall.

13. The method of claim 12, wherein analyzing the corpus of private electronic data comprises:

determining, by the processing device, whether the corpus of private electronic data was obtained from sources on an excluded list; and

if the corpus of data was obtained from sources on the excluded list, tagging, by the processing device, the corpus of private electronic data as denied for use outside the privacy firewall.

14. A method for implementing a privacy firewall to determine and provide non-private information from private electronic data to a remote computer located outside the privacy firewall, the method comprising:

analyzing, by a processing device, a corpus of private electronic data contained in a storage repository to identify a first one or more portions of the data comprising non-private information and a second one or more portions of the data comprising private information;

2015353713 09 Sep 2019

tagging, by the processing device, the first one or more portions of the data as allowed for use outside the privacy firewall;

determining, by the processing device, whether the second one or more portions of the data comprises non-private elements;

if the second one or more portions of the data comprises non-private elements:

extracting, by the processing device, the non-private elements, and

tagging, by the processing device, the non-private elements as information allowed for use outside the privacy firewall; and

providing, by the processing device, one or more of the first one or more portions of the data comprising non-private information and the non-private elements to the remote computer,

wherein determining whether the second one or more portions of the data comprises non-private elements comprises:

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that have appeared in other locations;

if the one or more sub-portions have appeared in other locations, tagging, by the processing device, the elements as non-private elements;

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that have appeared from a threshold number of distinct sources;

if the one or more elements have appeared from at least the threshold number of distinct sources, tagging, by the processing device, the one or more elements as non-private elements;

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that contain information that is already public knowledge;

if the one or more elements contain information that is already public knowledge, tagging, by the processing device, the one or more sub-portions as non-private elements;

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that are complex such that the one or elements are claimable as proprietary information;

if the one or more elements are complex, tagging, by the processing device, the one or more elements as denied for use outside the privacy firewall;

2015353713 09 Sep 2019

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that have recognizable sequences;

if the one or more elements have recognizable sequences, tagging, by the processing device, the one or more elements as denied for use outside the privacy firewall;

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that have precise timestamps; and

if the one or more elements have precise timestamps, tagging, by the processing device, the one or more elements as denied for use outside the privacy firewall.

15. The method of claim 12, wherein determining whether the second one or more portions of the data comprises non-private elements comprises:

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that have a timestamp that is below a granularity threshold; and

if the timestamp is below the granularity threshold:

tune, by the processing device, the one or more elements to a time range above the granularity threshold, and

tagging, by the processing device, the one or more tuned elements as non-private elements.

16. The method of claim 12, wherein determining whether the second one or more portions of the data comprises non-private elements comprises:

determining, by the processing device, whether the second one or more portions of the data comprises one or more elements that have a geographic location that is below a granularity threshold; and

if the geographic location is below the granularity threshold:

tune, by the processing device, the one or more elements to a geographic location above the granularity threshold, and

tagging, by the processing device, the one or more tuned elements as non-private elements.

2015353713 09 Sep 2019

17. The method of claim 12, further comprising:

providing, by the processing device, a response to the query to a user via a user interface, wherein the response to the query contains information obtained from one or more of the first one or more portions of the data comprising non-private information and the non-private elements.

18. A system for providing non-private information from private electronic data through a privacy firewall to a remote computer located outside the privacy firewall, the system comprising:

a data storage repository behind the privacy firewall, the data storage repository comprising a corpus of private electronic data that has been tagged as non-private information, private information not to be distributed, non-private elements of private information to be distributed, or depersonalized private information, wherein the non-private elements are portions of the private information or additional data that is derived from the private information that are not private;

a processing device; and

a non-transitory, processor-readable storage medium, wherein the non-transitory, processor-readable storage medium comprises one or more programming instructions that, when executed, cause the processing device to:

receive a search string from the remote computer located outside the privacy firewall, wherein the search string comprises a question;

search the data storage repository for one or more portions of the private electronic data that correspond to the search string;

if the one or more portions of the electronic data contain the non-private information, non-private elements of the private information to be distributed, or the depersonalized private information, provide a response to the remote computer, wherein the response contains information contained within the one or more portions of the electronic data contain the non-private information, the non-private elements of the private information to be distributed, or the depersonalized private information; and

tagging a remainder of the private information that is not the non-private information, non-private elements of the private information to be distributed, or the depersonalized private information as denied for use.

2015353713 09 Sep 2019

LEXISNEXIS A DIVISION OF REED ELSEVIER INC.

WATERMARK INTELLECTUAL PROPERTY PTY LTD

P42862AU00

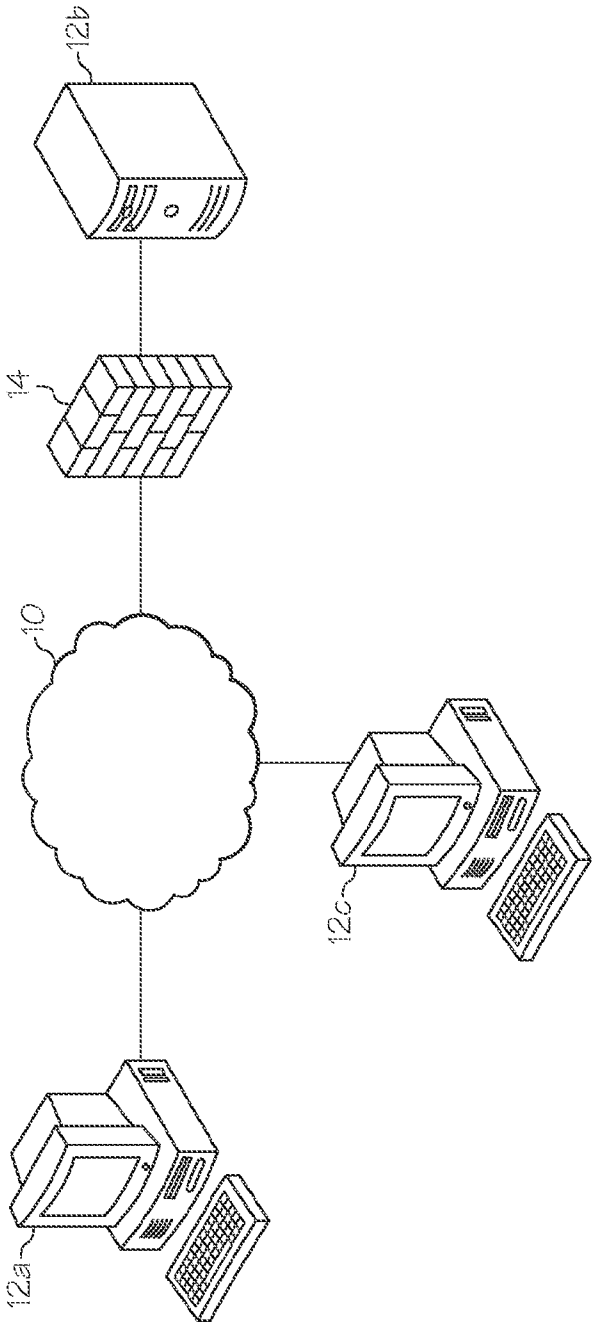

FIG. 1

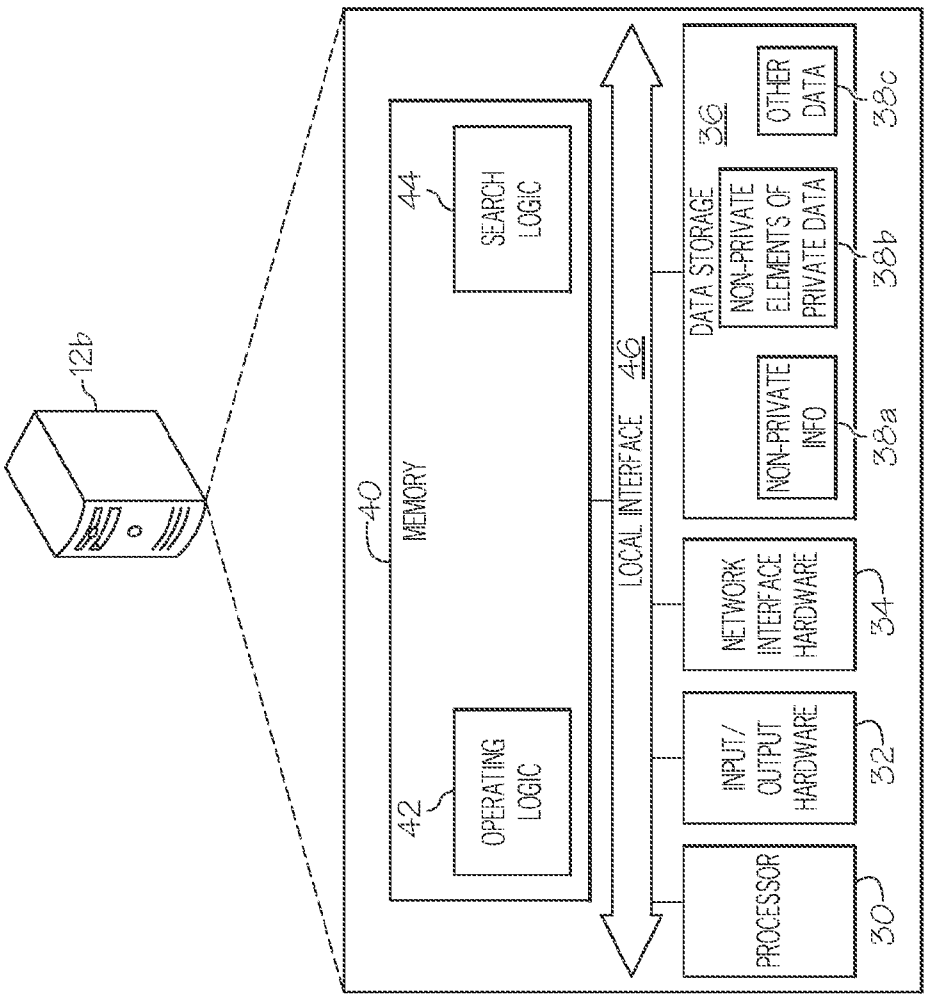
12b
40
MEMORY
42
OPERATING LOGIC
44
SEARCH LOGIC
LOCAL INTERFACE 46
DATA STORAGE 36
NON-PRIVATE INFO
38a
NON-PRIVATE ELEMENTS OF PRIVATE DATA
38b
OTHER DATA
38c
NETWORK INTERFACE HARDWARE
34
INPUT / OUTPUT HARDWARE
32
PROCESSOR
30

FIG. 2

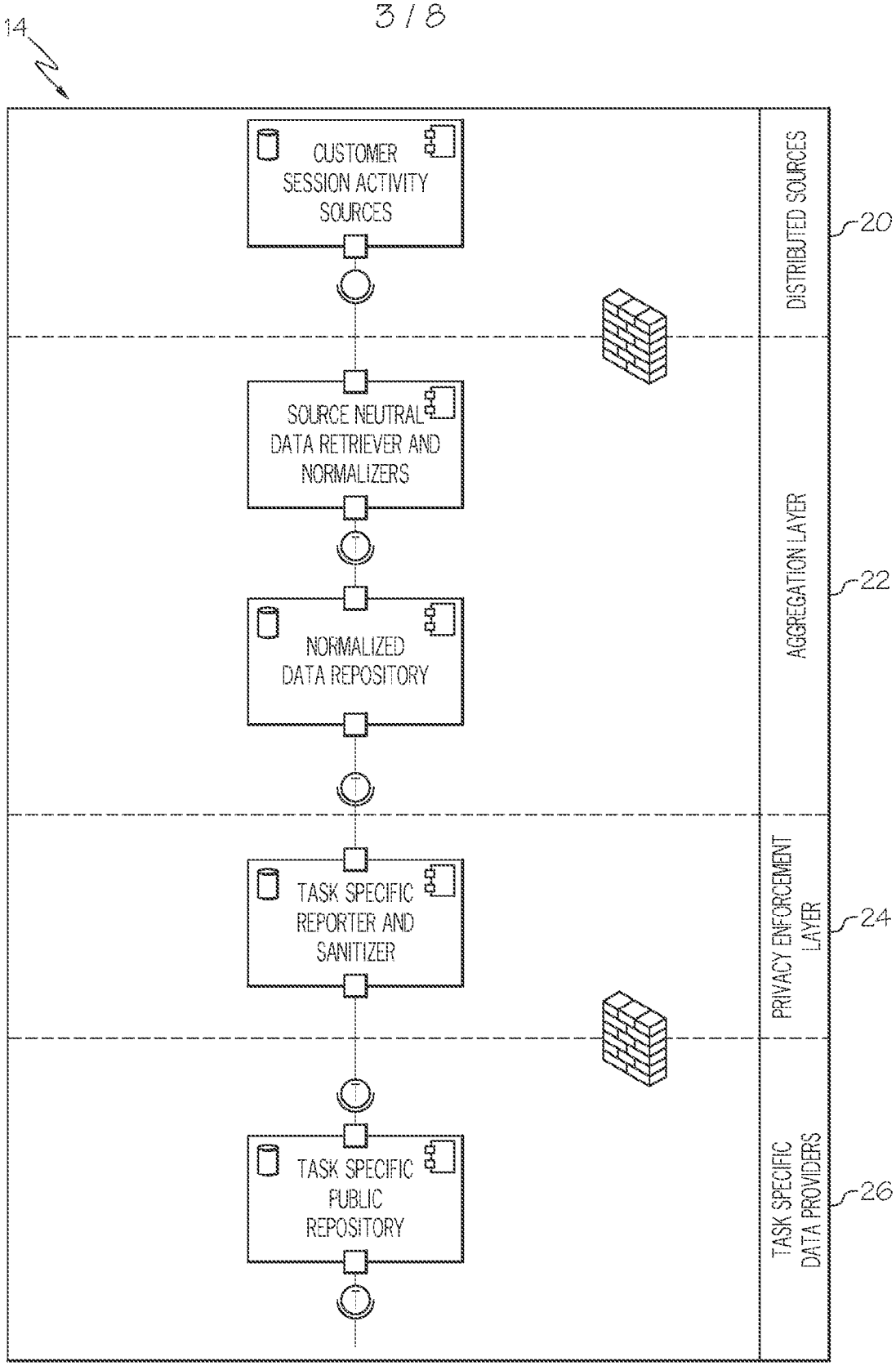
14
CUSTOMER SESSION ACTIVITY SOURCES
SOURCE NEUTRAL DATA RETRIEVER AND NORMALIZERS
NORMALIZED DATA REPOSITORY
TASK SPECIFIC REPORTER AND SANITIZER
TASK SPECIFIC PUBLIC REPOSITORY
DISTRIBUTED SOURCES
20
AGGREGATION LAYER
22
PRIVACY ENFORCEMENT LAYER
24
TASK SPECIFIC DATA PROVIDERS
26

FIG. 3

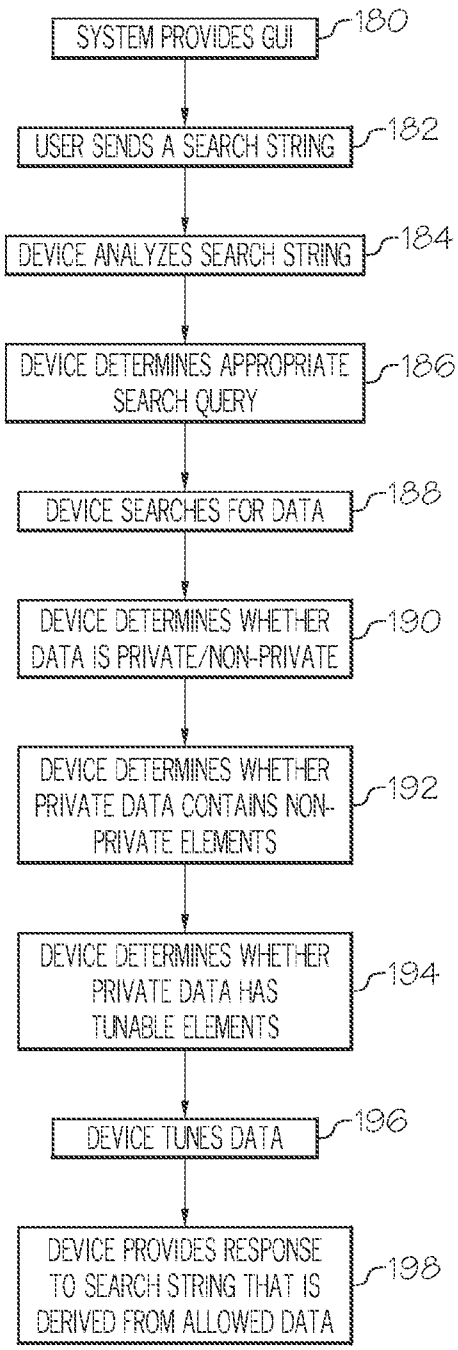
SYSTEM PROVIDES GUI
180
USER SENDS A SEARCH STRING
182
DEVICE ANALYZES SEARCH STRING
184
DEVICE DETERMINES APPROPRIATE SEARCH QUERY
186
DEVICE SEARCHES FOR DATA
188
DEVICE DETERMINES WHETHER DATA IS PRIVATE/NON-PRIVATE
190
DEVICE DETERMINES WHETHER PRIVATE DATA CONTAINS NON-PRIVATE ELEMENTS
192
DEVICE DETERMINES WHETHER PRIVATE DATA HAS TUNABLE ELEMENTS
194
DEVICE TUNES DATA
196
DEVICE PROVIDES RESPONSE TO SEARCH STRING THAT IS DERIVED FROM ALLOWED DATA
198

FIG. 4

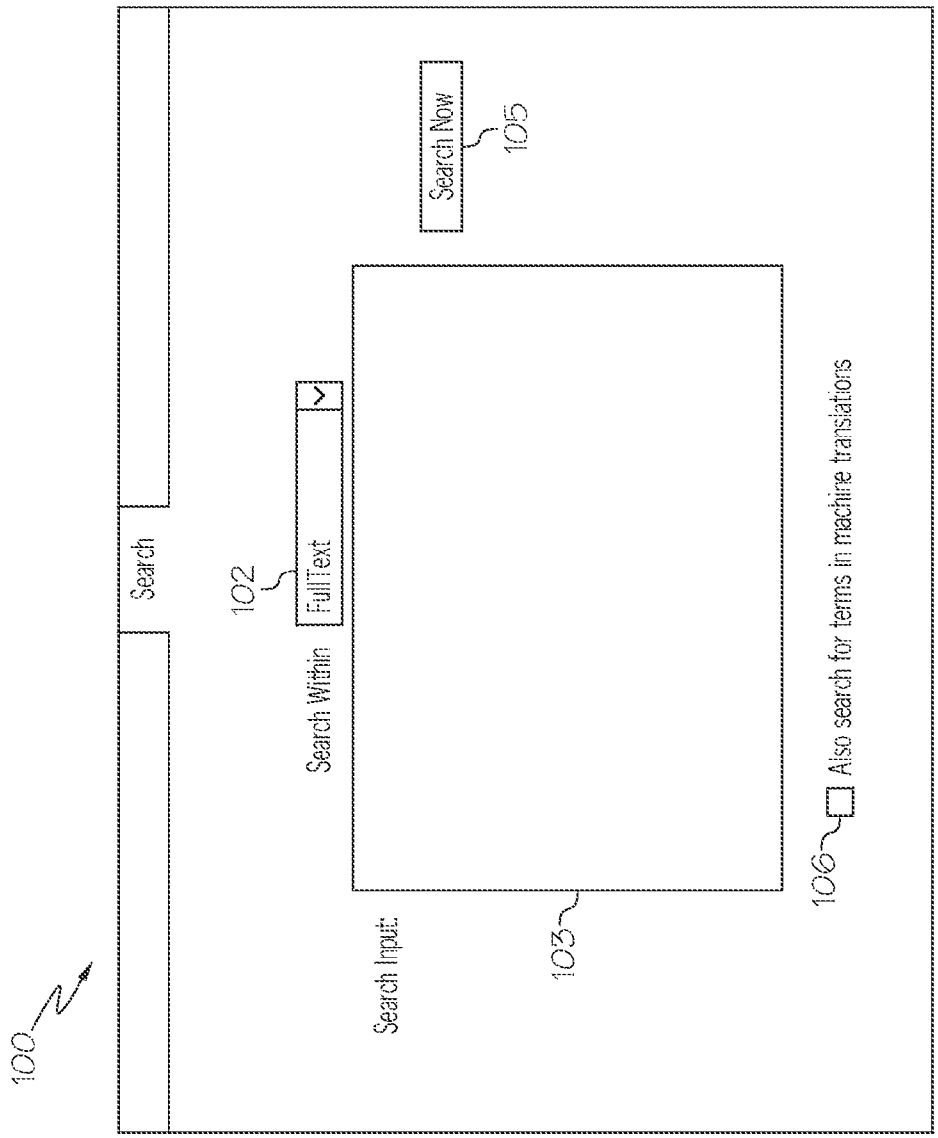
100
Search
Search Within
Full Text
102
Search Now
105
Search Input
103
Also search for terms in machine translations
106

FIG. 5

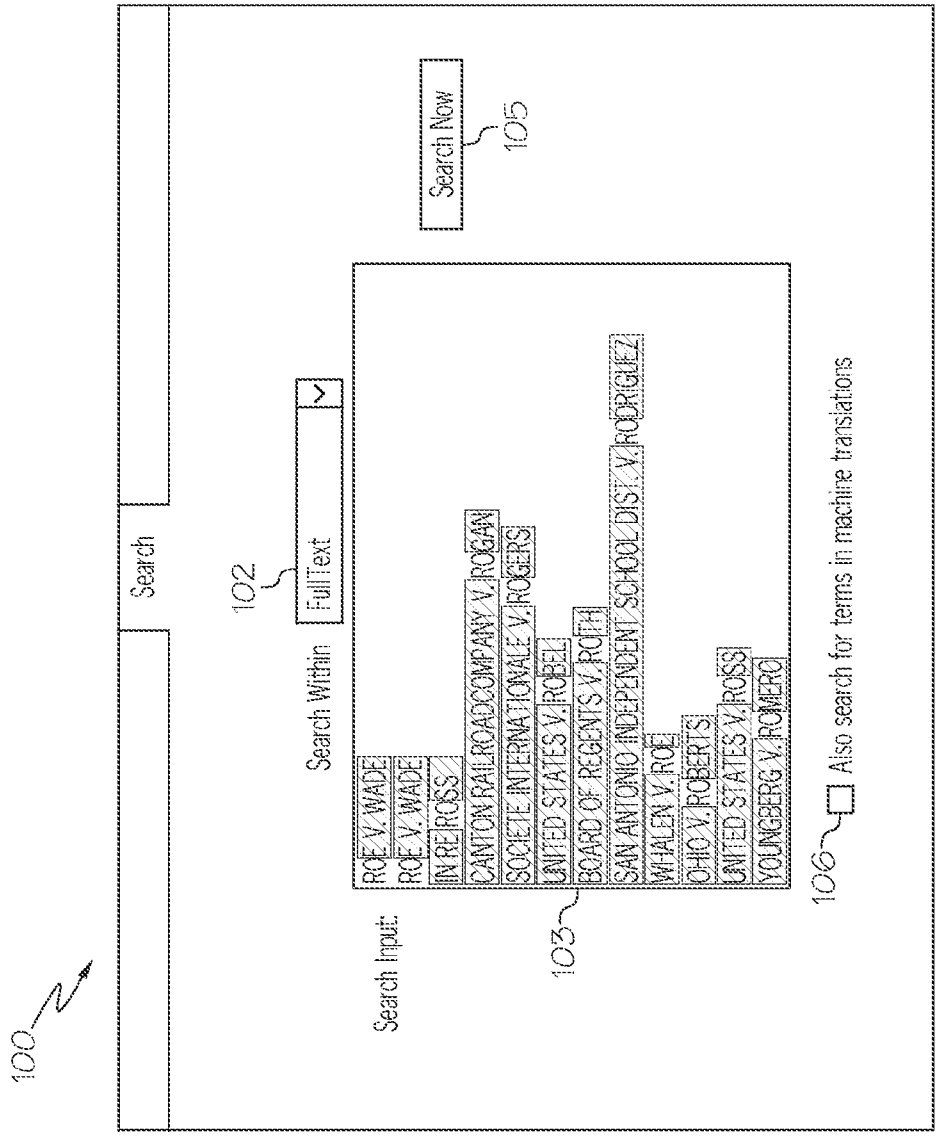
100
Search
Search Within
Full Text
102
Search Now
105
Search Input:
ROE V. WADE
ROE V. WADE
IN RE ROSS
CANTON RAILROAD COMPANY V. ROGAN
SOCIETE INTERNATIONALE V. ROGERS
UNITED STATES V. ROBEL
BOARD OF REGENTS V. ROTH
SAN ANTONIO INDEPENDENT SCHOOL DIST. V. RODRIGUEZ
WHALEN V. ROE
OHIO V. ROBERTS
UNITED STATES V. ROSS
YOUNGBERG V. ROMERO
103
106
Also search for terms in machine translations

FIG. 6

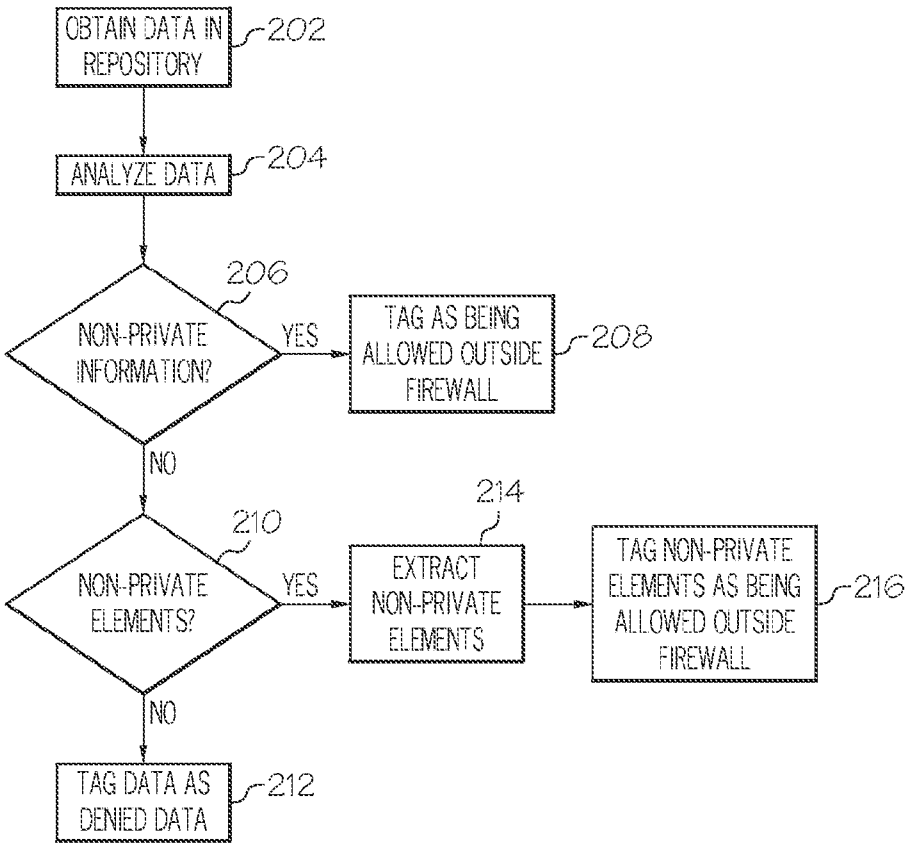
OBTAIN DATA IN REPOSITORY
202
ANALYZE DATA
204
206
NON-PRIVATE INFORMATION?
YES
TAG AS BEING ALLOWED OUTSIDE FIREWALL
208
NO
210
214
NON-PRIVATE ELEMENTS?
YES
EXTRACT NON-PRIVATE ELEMENTS
TAG NON-PRIVATE ELEMENTS AS BEING ALLOWED OUTSIDE FIREWALL
216
NO
TAG DATA AS DENIED DATA
212

FIG. 7

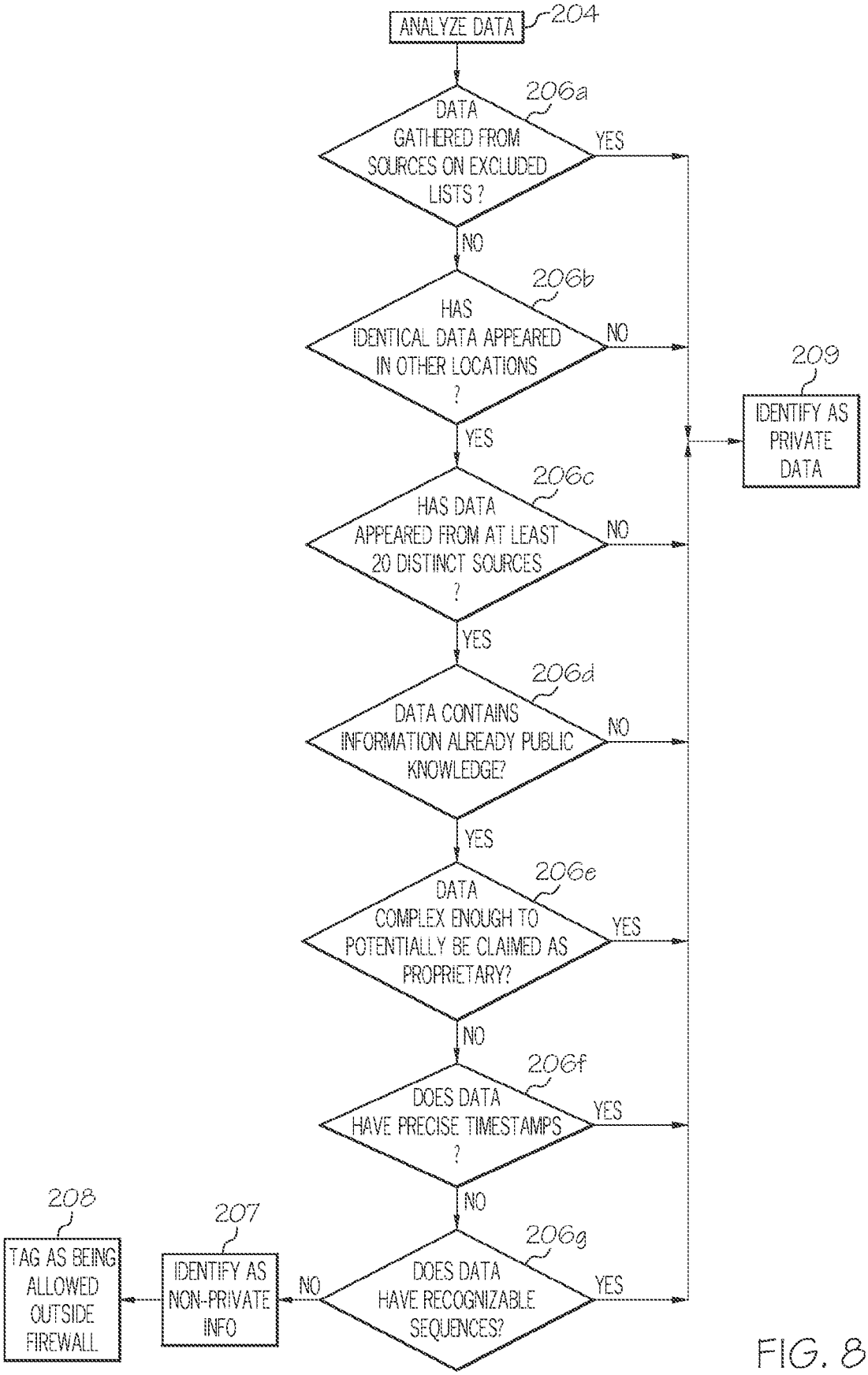
ANALYZE DATA
204
DATA GATHERED FROM SOURCES ON EXCLUDED LISTS ?
206a
YES
NO
HAS IDENTICAL DATA APPEARED IN OTHER LOCATIONS ?
206b
NO
YES
209
IDENTIFY AS PRIVATE DATA
HAS DATA APPEARED FROM AT LEAST 20 DISTINCT SOURCES ?
206c
NO
YES
DATA CONTAINS INFORMATION ALREADY PUBLIC KNOWLEDGE?
206d
NO
YES
DATA COMPLEX ENOUGH TO POTENTIALLY BE CLAIMED AS PROPRIETARY?
206e
YES
NO
DOES DATA HAVE PRECISE TIMESTAMPS ?
206f
YES
NO
DOES DATA HAVE RECOGNIZABLE SEQUENCES?
206g
YES
NO
IDENTIFY AS NON-PRIVATE INFO
207
TAG AS BEING ALLOWED OUTSIDE FIREWALL
208

FIG. 8