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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 April 2021 (22.04.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/076921 A1

(51) International Patent Classification:
G06F 7/04 (2006.01)

(21) International Application Number:
PCT/US2020/056026

(22) International Filing Date:
16 October 2020 (16.10.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/923,377	18 October 2019 (18.10.2019)	US
16/795,570	19 February 2020 (19.02.2020)	US
17/071,861	15 October 2020 (15.10.2020)	US

(71) Applicant: **ASG TECHNOLOGIES GROUP, INC. DBA ASG TECHNOLOGIES** [US/US]; 708 Goodlette Road North, Naples, Florida 34102 (US).

(72) Inventors: **MORESMAU, Jean-Philippe**; c/o ASG Technologies, 708 Goodlette Road North, Naples, Florida 34102 (US). **MACNEILL, Marcus**; c/o ASG Technologies, 708 Goodlette Road North, Naples, Florida 34102 (US).

(74) Agent: **GOEL, Sonia** et al.; Carr & Ferrell LLP, 120 Constitution Drive, Menlo Park, California 94025 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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).

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

(54) Title: USE OF MULTI-FACETED TRUST SCORES FOR DECISION MAKING, ACTION TRIGGERING, AND DATA ANALYSIS AND INTERPRETATION

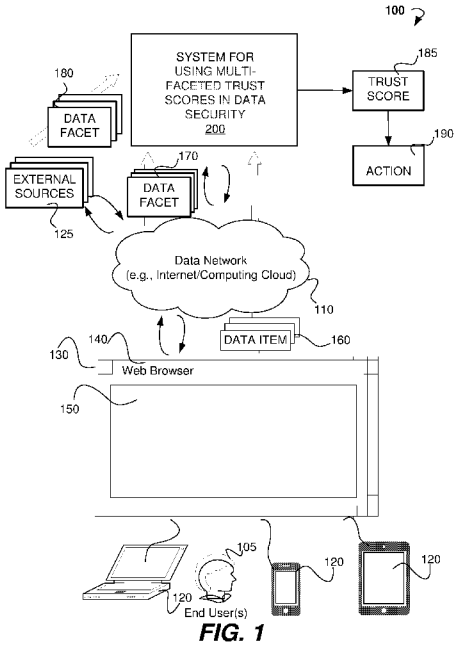

FIG. 1

(57) Abstract: Provided are methods and systems for using multi-faceted trust scores in data security. A method may commence with determine a plurality of data facets for a data item. The method may further include determining a plurality of parameters and weights corresponding to the plurality of data facets associated with the data item. Upon determining the plurality of parameters and weights, at least one trust score for the data item may be calculated based on the plurality of parameters and weights. The method may further include selectively taking an action based on the at least one trust score.

WO 2021/076921 A1

USE OF MULTI-FACETED TRUST SCORES FOR DECISION MAKING, ACTION TRIGGERING, AND DATA ANALYSIS AND INTERPRETATION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority of U.S. Provisional Patent Application No. 62/923,377 filed on October 18, 2019, entitled "Multi-Faceted Trust System," and is a Continuation-in-part of, and claims the priority benefit of, U.S. Patent Application No. 16/795,570 filed on February 19, 2020, entitled "Multi-Faceted Trust System," which are incorporated herein by reference in their entirety.

FIELD

[0002] This application relates generally to data processing and, more specifically, to systems and methods for using multi-faceted trust scores for decision making, action triggering, and data analysis and interpretation.

BACKGROUND

[0003] Approaches to understanding data may include defensive data strategies, such as regulatory compliance, and offensive data strategies, such as self-service data analytics. However, these approaches may be inefficient in some situations or even impossible to apply, e.g., when combining internal and external data, in driving additional insights into data and making decisions based on the data.

SUMMARY

[0004] This summary is provided to introduce a selection of concepts in a simplified form that are further described in the Detailed Description below. This summary is not intended to identify key features or essential features of the claimed subject matter, nor

is it intended to be used as an aid in determining the scope of the claimed subject matter.

[0005] Provided are methods and systems for using multi-faceted trust scores in data security. In some embodiments, a method for using multi-faceted trust scores in data security may commence with determining a plurality of data facets for a data item. The method may further include determining a plurality of parameters and weights corresponding to the plurality of data facets associated with the data item. Upon determining the plurality of parameters and weights, at least one trust score for the data item may be calculated based on the plurality of parameters and weights. The method may further include selectively taking an action based on the at least one trust score.

[0006] In some example embodiments, a system for using multi-faceted trust scores in data security may include a data analyzing unit, a score calculation unit, and a processing unit. The data analyzing unit may be configured to determine a plurality of data facets for a data item. The data analyzing unit may be further configured to determine a plurality of parameters and weights corresponding to the plurality of data facets associated with the data item. The score calculation unit may be configured to calculate at least one trust score for the data item based on the plurality of parameters and weights. The processing unit may be configured to selectively take an action based on the at least trust score.

[0007] Additional objects, advantages, and novel features will be set forth in part in the detailed description section of this disclosure, which follows, and in part will become apparent to those skilled in the art upon examination of this specification and the accompanying drawings or may be learned by production or operation of the example embodiments. The objects and advantages of the concepts may be realized and attained by means of the methodologies, instrumentalities, and combinations particularly pointed out in the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] Exemplary embodiments are illustrated by way of example and not limitation in the figures of the accompanying drawings, in which like references indicate similar elements.

[0009] FIG. 1 illustrates an environment within which methods and systems for using multi-faceted trust scores in data security can be implemented, according to an example embodiment.

[0010] FIG. 2 is a block diagram illustrating a system for using multi-faceted trust scores in data security, according to an example embodiment.

[0011] FIG. 3 is a flow diagram illustrating a method for using multi-faceted trust scores in data security, according to an example embodiment.

[0012] FIG. 4 is a schematic diagram showing multiple use cases for using multi-faceted trust scores in data security, according to an example embodiment.

[0013] FIG. 5 is a schematic diagram showing a widget for calculating and displaying a trust score, according to an example embodiment.

[0014] FIG. 6 is a computing system that can be used to implement a method for using multi-faceted trust scores in data security, according to an example embodiment.

DETAILED DESCRIPTION

[0015] The following detailed description includes references to the accompanying drawings, which form a part of the detailed description. The drawings show illustrations in accordance with exemplary embodiments. These exemplary embodiments, which are also referred to herein as "examples," are described in enough detail to enable those skilled in the art to practice the present subject matter. The embodiments can be combined, and other embodiments can be formed, by introducing structural and logical changes without departing from the scope of what is claimed. The following detailed description is, therefore, not to be taken in a limiting sense and the scope is defined by the appended claims and their equivalents.

[0016] In this document, the terms "a" or "an" are used, as is common in patent documents, to include one or more than one. In this document, the term "or" is used to refer to a nonexclusive "or," such that "A or B" includes "A but not B," "B but not A," and "A and B," unless otherwise indicated. Furthermore, all publications, patents, and patent documents referred to in this document are incorporated by reference herein in their entirety, as though individually incorporated by reference. In the event of inconsistent usages between this document and those documents so incorporated by reference, the usage in the incorporated reference(s) should be considered supplementary to that of this document; for irreconcilable inconsistencies, the usage in this document controls.

[0017] The present disclosure relates to systems and methods for using multi-faceted trust scores in data security. The system of the present disclosure provides uses of multi-faceted trust scores of data for decision making, action triggering, and data analysis and interpretation. The multi-faceted trust score of data may be provided for more comprehensive understanding of data and may be determined using a dynamic trust model including multiple facets. The system of the present disclosure can be used to analyze critical data components or other data items and determine a degree to

which the data can be trusted by the users. The degree to which the data can be trusted can be represented by a trust score of data. The trust score can be calculated based on several factors, which are referred to herein as data facets or facets. More specifically, each of the facets includes a characteristic of a data item (also referred to herein a metadata item). The trust score is also referred to as a "score" herein. The calculated trust score can be then used by the system to trigger actions, make decisions, perform data analysis, data interpretation, and so forth.

[0018] The method for using multi-faceted trust scores in data security may commence with receiving data including a plurality of data items and determining a plurality of data facets for each data item. The data facets may include objective data facets (metadata, data, business information), subjective data facets (ratings of a user), and synthetic data facets (calculated from heuristics on lineage). All data facets can be evaluated and scored. The data facets contribute to the final trust score based on their relative weights. The method may further include determining a plurality of parameters and weights corresponding to the plurality of data facets associated with the data items. In an example embodiment, the weight of a data facet can be entered by an operator based on empirical observations or established using a predetermined technique based on historical trustworthiness associated with the data facet. In some example embodiments, the parameters and weights of the facets can be determined by the system based on historical data associated with data items or facets. Upon determining the facets and corresponding parameters and weights, a trust score for each data items can be calculated based on the parameter and the weight associated with each of the facets. In an example embodiment, a trust score of the data may be calculated based on the trust scores calculated for each of the data items.

[0019] The method can further include selectively taking an action based on the trust score calculated for the data items. The action may include sending a notification to a user, triggering an automated action starting an intelligent business process

management workflow, integrating the trust score into a decision table, creating and sending a report, and so forth.

[0020] Referring now to the drawings, FIG. 1 illustrates an environment 100 within which systems and methods for using multi-faceted trust scores in data security can be implemented. The environment 100 may include a data network 110 (e.g., an Internet or a computing cloud), end user(s) 105, client device(s) 120 associated with the end user 105, external sources 125, and a system 200 for using multi-faceted trust scores in data security. Client device(s) 120 may include a personal computer (PC), a desktop computer, a laptop, a smartphone, a tablet, and so forth.

[0021] The client device 120 may have a user interface 130. Furthermore, a web browser 140 may be running on the client device 120 and displayed via the user interface 130. The web browser 140 may communicate with the system 200 via the data network 110.

[0022] The data network 110 may include the Internet or any other network capable of communicating data between devices. Suitable networks may include or interface with any one or more of, for instance, a local intranet, a corporate data network, a data center network, a home data network, a Personal Area Network, a Local Area Network (LAN), a Wide Area Network (WAN), a Metropolitan Area Network, a virtual private network, a storage area network, a frame relay connection, an Advanced Intelligent Network connection, a synchronous optical network connection, a digital T1, T3, E1 or E3 line, Digital Data Service connection, Digital Subscriber Line connection, an Ethernet connection, an Integrated Services Digital Network line, a dial-up port such as a V.90, V.34 or V.34bis analog modem connection, a cable modem, an Asynchronous Transfer Mode connection, or a Fiber Distributed Data Interface or Copper Distributed Data Interface connection. Furthermore, communications may also include links to any of a variety of wireless networks, including Wireless Application Protocol, General Packet Radio Service, Global System for Mobile Communication, Code Division Multiple

Access or Time Division Multiple Access, cellular phone networks, Global Positioning System, cellular digital packet data, Research in Motion, Limited duplex paging network, Bluetooth radio, or an IEEE 802.11-based radio frequency network. The data network can further include or interface with any one or more of a Recommended Standard 232 (RS-232) serial connection, an IEEE-1394 (FireWire) connection, a Fiber Channel connection, an IrDA (infrared) port, a Small Computer Systems Interface connection, a Universal Serial Bus (USB) connection or other wired or wireless, digital or analog interface or connection, mesh or Digi® networking.

[0023] The web browser 140 can render a web page associated with the system 200 which end user(s) 105 can use to calculate a trust score and take an action based on the calculated trust score. The web browser 140 can establish a communication channel with the system 200 and generate and render virtual screens based on data received from the system 200.

[0024] The system 200 may receive or collect data items 160 for which a trust score needs to be calculated. The data items 160 may be provided by the end user(s) 105 using the client device 120 or collected by the system 200 automatically. Upon receiving or collecting the data items 160, the system 200 may determine data facets 170 associated with the data items 160. The data facets 170 may include parameters of the data items 160. In some embodiments, the system 200 may also receive further data facets 180 related to the data items 160 from the external sources 125. Based on the data facets 170 and 180, the system 200 may calculate the trust score 185. The system 200 may render the results of calculation by the web browser 140 and display the trust score 185 to the end user(s) 105. Based on the trust score 185 determined for the data items 160, the system 200 may take an action 190. In an example embodiment, the action 190 may include determining whether the trust score 185 exceeds a predetermined threshold to decide whether the data including the data items 160 can be trusted. The action 190 may also include automatically starting an intelligent business process

management workflow, integrating the trust score into a decision table, creating a report based on the data items 160, and so forth.

[0025] FIG. 2 is a block diagram illustrating a system 200 for using multi-faceted trust scores in data security, according to an example embodiment. The system 200 may include a data analyzing unit 220, a score calculation unit 230, a processing unit 240, and, optionally, a data collection unit 210. In an example embodiment, the operations performed by each of the data collection unit 210, the data analyzing unit 220, the score calculation unit 230, and the processing unit 240 may be performed by a processor and a memory for storing instructions executable by the processor. Example one or more processors are shown in FIG. 6 as one or more processors 610. The operations performed by each of the data collection unit 210, the data analyzing unit 220, the score calculation unit 230, and the processing unit 240 of the system 200 are described in detail below with reference to FIG. 3.

[0026] FIG. 3 shows a process flow diagram of a method 300 for using multi-faceted trust scores in data security, according to an example embodiment. In some embodiments, the operations may be combined, performed in parallel, or performed in a different order. The method 300 may also include additional or fewer operations than those illustrated. The method 300 may be performed by processing logic that may comprise hardware (e.g., decision making logic, dedicated logic, programmable logic, and microcode), software (such as software run on a general-purpose computer system or a dedicated machine), or a combination of both.

[0027] The method 300 may commence with determining a plurality of data facets for a data item by the data analyzing unit at operation 310. In some embodiments, prior to the determination of the plurality of data facets, data including a plurality of data items may be received by the data collection unit. The data item for which the plurality of data facets is determined may be one of the plurality of data items. The data may include any type of data that needs to be analyzed, for which a trust score is required to

be determined, and based on which an action needs to be taken. The data may be received from any source (for example, a database, a remote source, an online source, a cloud database, the system 200 itself, and so forth).

[0028] Each of the plurality of data facets may include a characteristic of the data item. For example, the plurality of data facets may include one or more of the following: data quality dimensions, criticality of the data item, governance of the data item (governed/not governed), a rating, a review, an issue, a proximity of the data item to a source (for example, how close the data item is to the source), an existence of a data lineage, a fact of scanning of the data item from an active source or a spreadsheet, a tag associated with the data item, a frequency of update, a frequency of use, a null value, usefulness of the data item for an intended purpose, and so forth. The data facets may include objective data facets (metadata, data, business information), subjective data facets (for example, based on ratings of a user), and synthetic data facets (calculated from heuristics based on lineage).

[0029] At operation 320, the data analyzing unit may determine a plurality of parameters and weights corresponding to the plurality of data facets associated with the data item. The parameters may include, for example, a value of the data facet, a 'yes/no' parameter, a percentage, a range, information entered by a user in an application, information provided in any way by the user, whether the data item was analyzed, whether the data item was scheduled for the analysis, frequency of analysis of the data item, whether the data item was analyzed manually by a user or automatically by the system, and so forth.

[0030] The method 300 may continue with calculating at least one trust score for the data item based on the plurality of parameters and weights by the score calculation unit at operation 330. In an example embodiment, the method 300 may optionally include calculating a plurality of trust scores associated with each data facet of the plurality of data facets to obtain a plurality of trust scores for the plurality of data facets. The

calculation of the plurality of trust scores may be based on the plurality of parameters and weights associated with each data facet of the plurality of data facets. In an example embodiment, the at least one trust score of the data item may be a sum of the plurality of trust scores associated with each data facet of the plurality of data facets. Specifically, the calculation of the at least one trust score may include summing all parameters of facets (related to the metadata item) multiplied by corresponding weights and dividing the summation by the number of facets related to the data item. The result can be a number between 0 and 1 or a percentage between 0% and 100%.

[0031] The trust score may be calculated for each of the plurality of data items of data received by the data collection unit. Thereafter, the method 300 may proceed with calculating, by the score calculation unit, a data trust score for the data by summarizing the trust score for each of the plurality of data items and dividing the summation by the number of data items related to the data.

[0032] In an example embodiment, the method 300 may optionally include selecting, from the plurality of data facets, one or more data facets contributing to the trust score for each of the data items. In this embodiment, the calculation of the trust score for each of the plurality of data items may be based on the parameter and the weight associated with the selected one or more data facets.

[0033] A decision table may be used to determine which data facets contribute to the trust score of the data item (i.e., to determine which data facets to use and which weights to assign to each data facet). The decision table can use a type of the data item (for example, tables and columns may be evaluated differently) and other criteria. The result of the decision is a list of selected data facets with parameters and weights. The data intelligence (DI) repository may then calculate each data facet based on the information the data facet contains, using the parameters provided. For example, the DI repository may check if a business term was associated with the data item by determining whether the data item is tagged by the tag of type "Business Glossary."

Other types of tags may also be referenced in the trust facet configuration. All trust scores can be then aggregated, taking into account the weights of each data facet (some data facets may contribute more than others to the final trust score). The final score and each individual data facet score can be stored alongside the data item. The DI repository may provide some trust configurations that can be readjusted based on specific needs. Other data facets can be added to the data facets provided by default.

[0034] At operation 340, the processing unit may selectively take an action based on the at least one trust score. Therefore, once the trust score is calculated, the trust score can be used as an input for a variety of use cases inside a data intelligence model (e.g., a machine learning an artificial intelligence node) itself or other products. Specifically, the trust score can trigger predetermined actions when it reaches specific thresholds for the data item and/or can be used in conjunction with other data intelligence information to provide additional information associated with the data.

[0035] In some example embodiments, the selective taking of the action may be further based on information external to the data item. For example, further data facets associated with the data item may be received from external sources. In these embodiments, the trust score can be also combined with (or constructed entirely from) information from sources outside of the data intelligence model in order to drive additional information into data and decisions beyond the data intelligence model.

[0036] In an example embodiment, the action may include determining that the at least one trust score has exceeded a predetermined threshold or determining that the at least one trust score is lower than the predetermined threshold. The trust score may be expressed as a percentage. For example, the trust score that reached 100% threshold means "complete trust," while the trust score is at 0% means "no trust at all." The trust score and how each data facet has contributed to the trust score can be visualized in order to evaluate why the trust score is what it is and what has caused the trust score to change over time. Predetermined thresholds can be associated with different levels of

trust for more intuitive visualization; for example, a trust score that has exceeded 90% threshold can be shown in green, a trust score that is under the 50% threshold can be shown in red, and so forth. Data facet visualization may also be indicative of whether a data facet is objective (such as a metadata field value), subjective (such as user ratings), or synthetic (for example, a facet calculated from heuristics on lineage).

[0037] In an example embodiment, the action may include one or more of the following: sending a notification to at least one user, performing an automated action starting an intelligent business process management workflow, integrating the at least one trust score into a decision table, creating a report, and so forth.

[0038] In an example embodiment, the at least one trust score of the data item may be combined with at least one lineage associated with the data item. The at least one lineage may be based on evaluation of data provenance of the data item. In this embodiment, the selection of the action may be based on the combination of the at least trust score and the at least one lineage.

[0039] The method 300 may further include determining that the data item has been changed. The change may include at least one of the following: re-analysis of the data item, manual tagging of the data item, a rating assigned to the data item, a trigger initiated by a user, and so forth. Based on this determination, the at least one trust score can be recalculated for the data item.

[0040] In an example embodiment, facets that contribute to the trust score can be defined in policies. Specifically, a policy type of "enrichment" can be defined and used for any data that can contribute additional data. The facets may include a type of the data item, a catalog in which the data item is placed, and so forth. More facets can be added if additional fine-grained calculations are needed. The result of setting the policy is a list of "Trust Facets." A facet can define what data item needs to be considered and the relative weight of the facet in the overall calculation. Upon setting the policies, the

trust score can be calculated by a backend service that reacts to events. The trust score and an explanation as to what facets contributed to the trust score can be displayed.

[0041] The facets contributing to the trust score may include the following data facets: quality, ratings average, ratings count, tags, linked items, and so forth. The weight of each data facet may be determined or set in the range between 0 and 1. The quality may be calculated by a quality process (which can be driven by a configuration). The ratings average is an average of ratings given to the data item, e.g., when ratings go between 1 and 5, 3 is average. A rating of 1 results in a score of 0, a rating of 5 results in a score of 1, and a rating of 3 results in a score of 0.5. The ratings count is a minimum value for the count of ratings to count. If there are more, the score weighted by the count for each rating is calculated. Since ratings with low counts are filtered out, the ratings count can give a different result than the average. The tags can include determining whether the data item is tagged by any tag in the hierarchy of the given tag name. For example, any tag created from a glossary term may increase the trust score, meaning that if a data item is linked to a glossary term, the data item has been analyzed somewhat and is trusted more. The linked items may be taken and the average of their trust score can be calculated. For example, a trust score of a dataset may include the trust score of the data items referenced by the dataset, the score of a table may include the score of each column, and so forth.

[0042] FIG. 4 is a schematic diagram 400 showing multiple use cases for using multi-faceted trust scores in data security, according to an example embodiment. Upon receiving data 405 that include a plurality of data items, a trust scope 410 for the data 405 may be calculated. When the trust score 410 passes a predetermined threshold, a predetermined action 415 can be triggered. The predetermined action 415 may include sending a notification to specific users, triggering automated actions, such as starting an intelligent business process management workflow, generating a report, and so forth. For example, a user may generate a report based on data available in the system. Upon

reviewing the trust score for the data based on which the report was generated, the user may determine, based on the trust score, whether the data can be trusted and whether the report is based on the trusted data.

[0043] In an example embodiment, the action 415 may include providing a recommendation on how to increase the trust score 410 of the data 405. The recommendation may include performing additional analysis, linking the data 405 to specific datasets, adding additional data items to the data 405, scheduling an automatic periodical analysis of the data 405, and so forth.

[0044] In an example embodiment, the action 415 may include marking one or more data items of data 405 as critical. For example, when the trust score 410 of the data 405 or individual data items of the data 405 becomes lower than a predetermined threshold, the data 405 or individual data items of the data 405 may be marked as critical and a notification informing about critical data may be sent to a user. The user may perform an investigation of changes that lowered the trust score 410 of the data 405 or individual data items of the data 405.

[0045] In an example embodiment, the action 415 may include sending a notification to the user when the trust score 410 exceeds a value preselected by a user. For example, the user may want to use the data 405, but cannot use the data 405 at the moment because the data 405 have a low trust score. The user may set a value of the trust score 410, e.g., 60%, upon exceeding which by the data 405 the notification should be sent to the user.

[0046] In another example embodiments, the trust score 410 can be integrated into the decision making 420. The trust score 410 can become a factor in decision tables used by a policy engine to drive decisions inside and outside of the data intelligence model associated with the system 200. For example, the decision as to whether the data 405 can be used to create reports can be based on the trust score 410.

[0047] Further example decision can be made based on the trust score 410 including providing access to the data 405 to specific users based on the data process information. The data process information may include, among other things, the trust score 410 of the data 405. For example, a database table may be restricted from being shown to prospective customers if the trust score 410 of the database table is below a predetermined threshold, i.e., when the data in the database table are not reliable. Vice versa, if the trust score 410 of the database table exceeds the predetermined threshold, a decision to show the database table to customers may be taken because there is high probability that the data in the database table are accurate.

[0048] In a further example embodiment, a portion of data may be hidden from users based on privileges assigned to the users. For example, the privileges assigned to the users may be used to calculate the trust score of the data and to determine whether the data or a portion of the data needs to be hidden from a user having specific privileges. For example, customer names in a database table may be hidden from users having a low privilege in the system in which the database table is stored.

[0049] In a further example embodiment, the data 405 may include real-world data, e.g., customer information, which may not be allowed to be shown to users. Based on the trust score 410, a decision to convert at least a portion of the data 405 into metadata allowed to be shown to the users may be taken. For example, a customer name (real-world data) can be converted into a customer identification number (metadata), a total amount of money paid by a customer (real-world data) can be converted into a minimum value and a maximum value of the amount paid by the customer (metadata).

[0050] In a further example embodiment, the decision may be made based on whether the data 405 are collected by the system of the user from internal sources (e.g., inside an organization of the user) or whether at least a portion of the data is collected from outside sources (e.g., county statistic data). The data that include information

collected from outside sources may have a lower trust score than the data managed completely by the user.

[0051] In an example embodiment, the trust score 410 may be combined with lineage results shown as data lineage 425. A low-trust table (i.e., a table having a low trust score) may be connected via the data lineage with a high-trust table, which may result in a potential issue (e.g., when the first table needs to be trusted more or the second table needs to be trusted less) or an indication that the data lineage covers a "cleaning up" operation. An indication showing the trust score of tables and the data lineage of the table may be provided to the user for review.

[0052] In further example embodiment, a high-trust table (i.e., a table having a high trust score) may be connected via the data lineage with a low trust table, which may result in a potential issue (the first table needs to be trusted less or the second needs to be trusted more) or an indication that the data lineage covers integrating less trusted data or performs dubious calculations. These issue may be reviewed by a user.

[0053] Moreover, the data lineage may be used in the trust calculation itself: a table the data provenance of which is proven by the data lineage may have a higher trust score than a table the data provenance of which is unknown. Tables derived from highly trusted data may be more trusted unless the lineage transformation itself is not trusted. Tables contributing to highly trusted tables can be considered more trusted (or indicate that these tables need a higher level of scrutiny to ensure the downstream data should effectively be trusted).

[0054] In an example embodiment, the system 200 can be used to determine how much a set of data lineage information can be relied on, a confidence level related to the results representing all the places to which a data item may have flowed, a confidence level related to the fact that a source of information is source X (but not source Y), a confidence level related to the fact that a source is a reliable source, and so forth.

[0055] The data lineage may include information on how data in different parts of an organization (e.g., customer information, a database, employee information, and the like) relate to each other. For example, the data may include a trial table storing information about users that inquired about a product on a website associated with the organization. The data may further include a customer database storing information of customers that already bought a product. The lineage may show that some of the customers in the customer database bought the product after they asked about the product on the website and were contacted by a manager of the website. Therefore, even if the information of customers and prospective customers is stored in the same customer database, the lineage may show specific details as to where the customer information is received from. This lineage may be used, for example, when specific customers need to be selected and stored in an email campaign table.

[0056] In an example embodiment, a first table may store customer names. The lineage may show that the customer names were copied from a second table having a high trust score. Therefore, based on the lineage, the first table may also have the high trust score.

[0057] The lineage can show that data flow from source A to source B but the trust scores of source A and source B are different, which may be an indication that there is an issue with the data that needs to be investigated by the user. For example, a user may have a table with a high trust score and needs to add data to the table, and the lineage can show that a source from which the data needs to be added has a low trust score. In another example, the user may have data with a high trust score. An application may use these data and provide a result. The lineage can show that the application has a low trust score, e.g., because the user has no source code for the application and the operations performed by the application are unknown. Therefore, the result provided by the application may have a trust score lower that the trust score of the original data used by the application.

[0058] In another example embodiment, the data in a database may come from a highly trusted source and copied to a table for review by customers. However, the table may have a low trust score because the customers did not review the table or did not rate the table for some reasons. Based on the lineage showing that the data in the table have a highly trusted source, the trust score of the table for customers may be increased.

[0059] In another example embodiment, the trust score of the data may be based on ratings provided to the data by the users. The lineage may show that some users rate high everything they review. Therefore, the rating provided by these users may be of low trust. In this case, a weight assigned to ratings may be lowered. The lineage may further show that the data are received from five tables each having a specific trust score. The weight assigned to the trust scores of these five tables may be increased. The trust score of the data may be determined as an average of the trust scores of these five tables.

[0060] In further example embodiments, the trust score 410 can be used in record management 430. For example, the trust score 410 can be used to determine a confidence level of the fact that a data item is or is not personally identifying information, that a given alert (raised by any of products/applications) is not a false positive, that a robotic automation process can work correctly each time (e.g., if there is a web page position dependency, the risk that the process can be broken by someone changing the page is high), that a set of policies can work as intended (e.g., if there are multiple nested policies or policies specified by different people, the side effects resulting from the ways in which they combine can result in unintended consequences), a confidence level of a validation process (e.g., whether every figure in the quarterly report was validated, and what level of confidence can be associated with each of the sources and calculations that informed that calculation), a confidence level of the fact that a given document is a reliable source of information, that all personally identifying

information in a document (or all documents) is redacted, that a document was classified correctly, and so forth.

[0061] In terms of data mining, the trust score 410 can be used to determine whether content metadata has been extracted from a document, whether content metadata was extracted correctly, whether content metadata was classified correctly (e.g., that 1234565890 is really an account number and not an invoice number), and so forth.

[0062] In terms of record management 430, the trust score 410 can be also used to determine whether the content can be deleted, or whether long term/permanent retention of the content can be applied. In some embodiment, the trust score 410 can be applied to documents to determine whether a document has a trust score that exceeds a predetermined threshold.

[0063] FIG. 5 is a schematic diagram showing a widget 500 for calculating and displaying a trust score 505, according to an example embodiment. The user interface may be using a chart 510, such as an ngx-chart (Gauge), to show the trust score 505 graphically. Red/yellow/green colors 515 may be shown based on the trust score 505 (green if over 90%, yellow if over 60%, and red otherwise). The colors can be configurable. Upon hovering over the chart 510 by the user, corresponding explanations 520 can be shown to the user. In an example embodiment, a widget, such as the widget 500 showing the trust score 505 associated with the quality, can be shown in search results when the user search for specific data.

[0064] FIG. 6 illustrates an exemplary computing system 600 that may be used to implement embodiments described herein. The exemplary computing system 600 of FIG. 6 may include one or more processors 610 and memory 620. Memory 620 may store, in part, instructions and data for execution by the one or more processors 610. Memory 620 can store the executable code when the exemplary computing system 600 is in operation. The exemplary computing system 600 of FIG. 6 may further include a

mass storage 630, portable storage 640, one or more output devices 650, one or more input devices 660, a network interface 670, and one or more peripheral devices 680.

[0065] The components shown in FIG. 6 are depicted as being connected via a single bus 690. The components may be connected through one or more data transport means. The one or more processors 610 and memory 620 may be connected via a local microprocessor bus, and the mass storage 630, one or more peripheral devices 680, portable storage 640, and network interface 670 may be connected via one or more input/output buses.

[0066] Mass storage 630, which may be implemented with a magnetic disk drive or an optical disk drive, is a non-volatile storage device for storing data and instructions for use by a magnetic disk or an optical disk drive, which in turn may be used by one or more processors 610. Mass storage 630 can store the system software for implementing embodiments described herein for purposes of loading that software into memory 620.

[0067] Portable storage 640 may operate in conjunction with a portable non-volatile storage medium, such as a compact disk (CD) or digital video disc (DVD), to input and output data and code to and from the computing system 600 of FIG. 6. The system software for implementing embodiments described herein may be stored on such a portable medium and input to the computing system 600 via the portable storage 640.

[0068] One or more input devices 660 provide a portion of a user interface. The one or more input devices 660 may include an alphanumeric keypad, such as a keyboard, for inputting alphanumeric and other information, or a pointing device, such as a mouse, a trackball, a stylus, or cursor direction keys. Additionally, the computing system 600 as shown in FIG. 6 includes one or more output devices 650. Suitable one or more output devices 650 include speakers, printers, network interfaces, and monitors.

[0069] Network interface 670 can be utilized to communicate with external devices, external computing devices, servers, and networked systems via one or more communications networks such as one or more wired, wireless, or optical networks

including, for example, the Internet, an intranet, LAN, WAN, cellular phone networks (e.g., Global System for Mobile communications network, packet switching communications network, circuit switching communications network), Bluetooth radio, and an IEEE 802.11-based radio frequency network, among others. Network interface 670 may be a network interface card, such as an Ethernet card, optical transceiver, radio frequency transceiver, or any other type of device that can send and receive information. Other examples of such network interfaces may include Bluetooth®, 3G, 4G, and Wi-Fi® radios in mobile computing devices as well as a USB.

[0070] One or more peripheral devices 680 may include any type of computer support device to add additional functionality to the computing system. The one or more peripheral devices 680 may include a modem or a router.

[0071] The components contained in the exemplary computing system 600 of FIG. 6 are those typically found in computing systems that may be suitable for use with embodiments described herein and are intended to represent a broad category of such computer components that are well known in the art. Thus, the exemplary computing system 600 of FIG. 6 can be a PC, handheld computing device, telephone, mobile computing device, workstation, server, minicomputer, mainframe computer, or any other computing device. The computer can also include different bus configurations, networked platforms, multi-processor platforms, and so forth. Various operating systems (OS) can be used including UNIX, Linux, Windows, Macintosh OS, Palm OS, and other suitable operating systems.

[0072] Some of the above-described functions may be composed of instructions that are stored on storage media (e.g., computer-readable medium). The instructions may be retrieved and executed by the processor. Some examples of storage media are memory devices, tapes, disks, and the like. The instructions are operational when executed by the processor to direct the processor to operate in accord with the example

embodiments. Those skilled in the art are familiar with instructions, processor(s), and storage media.

[0073] It is noteworthy that any hardware platform suitable for performing the processing described herein is suitable for use with the example embodiments. The terms "computer-readable storage medium" and "computer-readable storage media" as used herein refer to any medium or media that participate in providing instructions to a central processing unit (CPU) for execution. Such media can take many forms, including, but not limited to, non-volatile media, volatile media, and transmission media. Non-volatile media include, for example, optical or magnetic disks, such as a fixed disk. Volatile media include dynamic memory, such as random access memory (RAM). Transmission media include coaxial cables, copper wire, and fiber optics, among others, including the wires that include one embodiment of a bus. Transmission media can also take the form of acoustic or light waves, such as those generated during radio frequency and infrared data communications. Common forms of computer-readable media include, for example, a floppy disk, a flexible disk, a hard disk, magnetic tape, any other magnetic medium, a CD-read-only memory (ROM) disk, DVD, any other optical medium, any other physical medium with patterns of marks or holes, a RAM, a PROM, an EPROM, an EEPROM, a FLASHEPROM, any other memory chip or cartridge, a carrier wave, or any other medium from which a computer can read.

[0074] Various forms of computer-readable media may be involved in carrying one or more sequences of one or more instructions to a CPU for execution. A bus carries the data to system RAM, from which a CPU retrieves and executes the instructions. The instructions received by system RAM can optionally be stored on a fixed disk either before or after execution by a CPU.

[0075] Thus, various embodiments of methods and systems for using multi-faceted trust scores in data security have been described. Although embodiments have been described with reference to specific example embodiments, it will be evident that

various modifications and changes can be made to these example embodiments without departing from the broader spirit and scope of the present application. Accordingly, the specification and drawings are to be regarded in an illustrative rather than a restrictive sense. There are many alternative ways of implementing the present technology. The disclosed examples are illustrative and not restrictive.

CLAIMS

What is claimed is:

1. A method for using multi-faceted trust scores in data security, the method comprising:

determining a plurality of data facets for a data item;

determining a plurality of parameters and weights corresponding to the plurality of data facets associated with the data item;

calculating at least one trust score for the data item based on the plurality of parameters and weights; and

selectively taking an action based on the at least one trust score.

2. The method of claim 1, further comprising calculating a plurality of trust scores associated with each data facet of the plurality of data facets to obtain a plurality of trust scores for the plurality of data facets, the calculation being based on the plurality of parameters and weights associated with each data facet of the plurality of data facets.

3. The method of claim 2, wherein the at least one trust score of the data item is a sum of the plurality of trust scores associated with each data facet of the plurality of data facets.

4. The method of claim 1, wherein the action includes one or more following:

determining that the at least one trust score exceeds a predetermined threshold; and

determining that the at least one trust score is lower than the predetermined threshold.

5. The method of claim 1, wherein the selectively taking the action is further based on information external to the data item.

6. The method of claim 1, wherein the action includes one or more of the following: notifying at least one user, an automated action starting an intelligent business process management workflow, integrating the at least one trust score into a decision table, and creating a report.

7. The method of claim 1, further comprising combining the at least one trust score of the data item with at least one lineage associated with the data item, the action being taken based on the at least trust score and the at least one lineage.

8. The method of claim 6, wherein the at least one lineage is based on evaluation of data provenance of the data item.

9. The method of claim 1, wherein each of the plurality of data facets includes a characteristic of the data item.

10. The method of claim 1, wherein the plurality of data facets include one of the following: data quality dimensions, criticality of the data item, governance of the data item, a rating, a review, an issue, a proximity of the data item to a source, an existence of a data lineage, a fact of scanning the data item from an active source or a spreadsheet, a tag associated with the data item, a frequency of update, a frequency of use, a null value, and usefulness of the data item for an intended purpose.

11. The method of claim 1, further comprising:

determining that the data item has been changed; and

based on the determination, recalculating the at least one trust score for the data item.

12. The method of claim 1, further comprising receiving data, the data including a plurality of data items, the plurality of data items including at least the data item;

calculating a trust score for each of the plurality of data items; and

calculating a data trust score by summarizing the trust score for each of the plurality of data items.

13. A system for using multi-faceted trust scores in data security, the system comprising:

a data analyzing unit configured to:

determine a plurality of data facets for a data item; and

determine a plurality of parameters and weights corresponding to the plurality of data facets associated with the data item; and

a score calculation unit configured to:

calculate at least one trust score for the data item based on the plurality of parameters and weights; and

a processing unit configured to:

selectively take an action based on the at least trust score.

14. The system of claim 13, further comprising a data collection unit configured to receive data, the data including a plurality of data items, the plurality of data items including at least the data item.

15. The system of claim 13, wherein the score calculation unit is further configured to calculate a plurality of trust scores associated with each data facet of the plurality of data facets to obtain a plurality of trust scores for the plurality of data facets, the calculation being based on the plurality of parameters and weights associated with each data facet of the plurality of data facets.

16. The system of claim 15, wherein the at least one trust score of the data item is a sum of the plurality of trust scores associated with each data facet of the plurality of data facets.

17. The system of claim 13, wherein the action includes one or more following:

determining that the at least one trust score exceeds a predetermined threshold; and

determining that the at least one trust score is lower than the predetermined threshold.

18. The system of claim 13, wherein the action includes one or more of the following: notifying at least one user, an automated action starting an intelligent business process management workflow, integrating the at least one trust score into a decision table, and creating a report.

19. The system of claim 13, wherein the plurality of data facets include one of the following: data quality dimensions, criticality of the data item, governance of the data item, a rating, a review, an issue, a proximity of the data item to a source, an existence of a data lineage, a fact of scanning the data item from an active source or a spreadsheet, a tag associated with the data item, a frequency of update, a frequency of use, a null value, and usefulness of the data item for an intended purpose.

20. A system for using multi-faceted trust scores in data security, the system comprising:

a data collection unit configured to receive data, the data including a plurality of data items, the plurality of data items including at least the data item;

a data analyzing unit configured to:

determine a plurality of data facets for a data item;

determine a plurality of parameters and weights corresponding to the plurality of data facets associated with the data item; and

a score calculation unit configured to:

calculate at least one trust score for the data item based on the plurality of parameters and weights; and

combine the at least one trust score of the data item with at least one lineage associated with the data item, wherein the at least one lineage is based on evaluation of data provenance of the data item; and

a processing unit configured to:

selectively take an action based on the at least trust score and the at least one lineage.

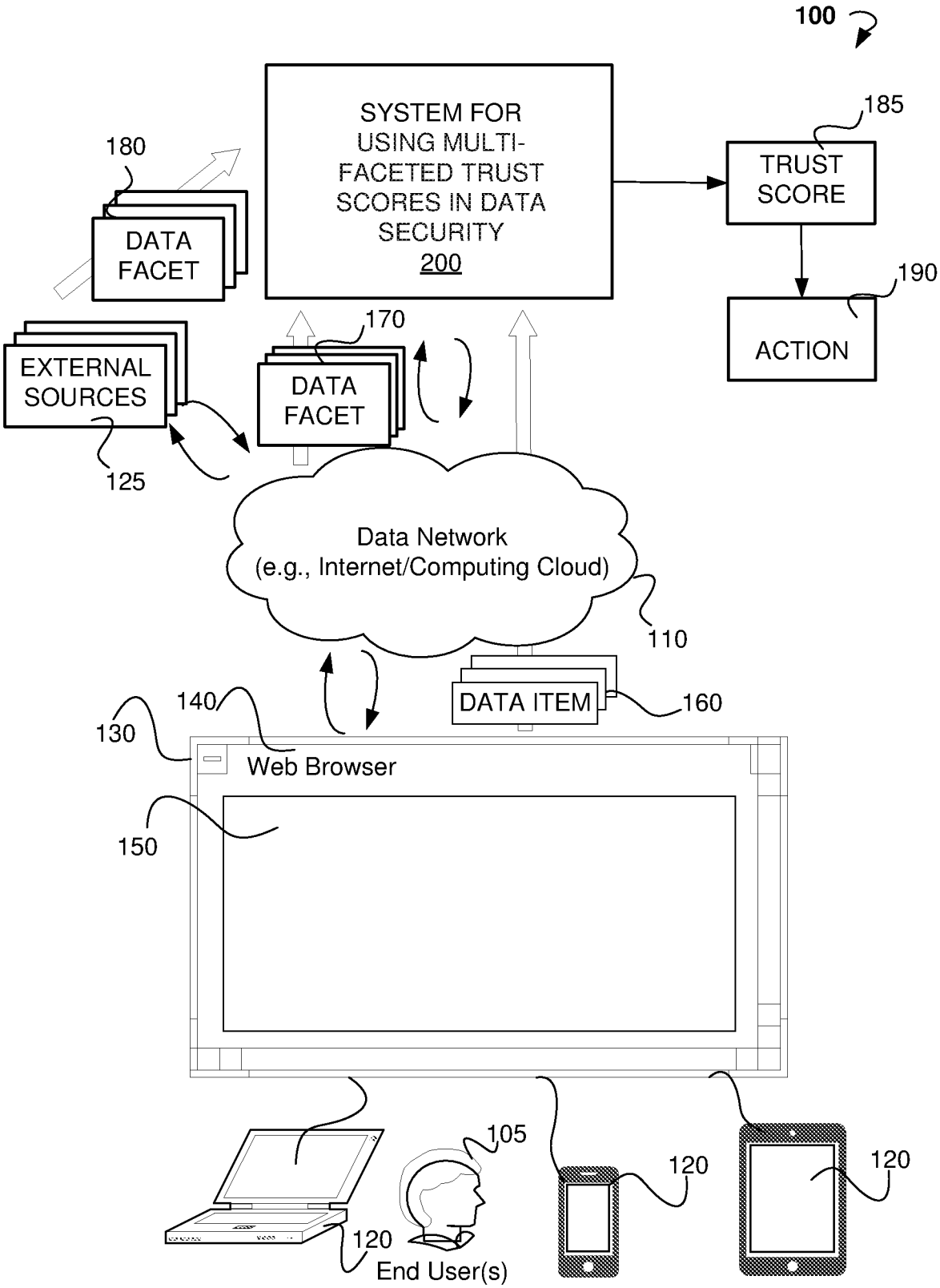
100
SYSTEM FOR USING MULTI-FACETED TRUST SCORES IN DATA SECURITY
200
185
TRUST SCORE
190
ACTION
180
DATA FACET
170
DATA FACET
EXTERNAL SOURCES
125
Data Network
(e.g., Internet/Computing Cloud)
110
DATA ITEM
160
140
130
Web Browser
150
105
120
120
120
End User(s)

FIG. 1

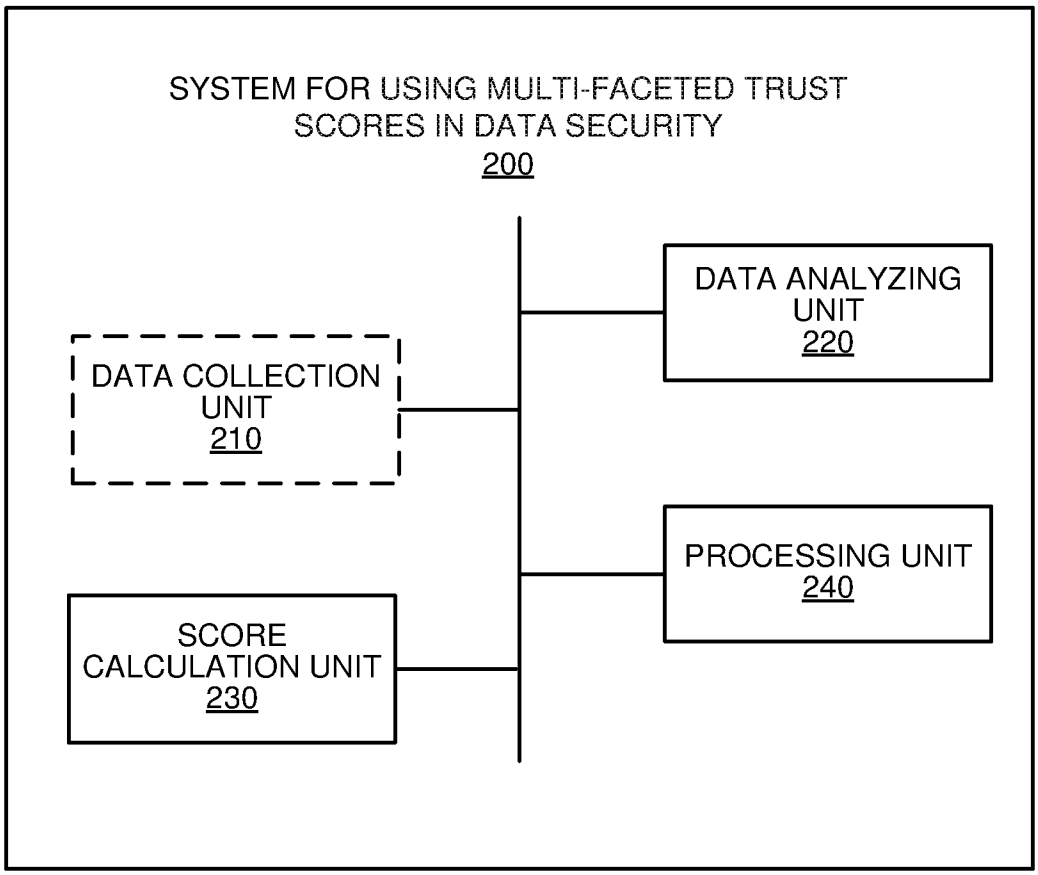
SYSTEM FOR USING MULTI-FACETED TRUST SCORES IN DATA SECURITY
200
DATA ANALYZING UNIT
220
DATA COLLECTION UNIT
210
PROCESSING UNIT
240
SCORE CALCULATION UNIT
230

FIG. 2

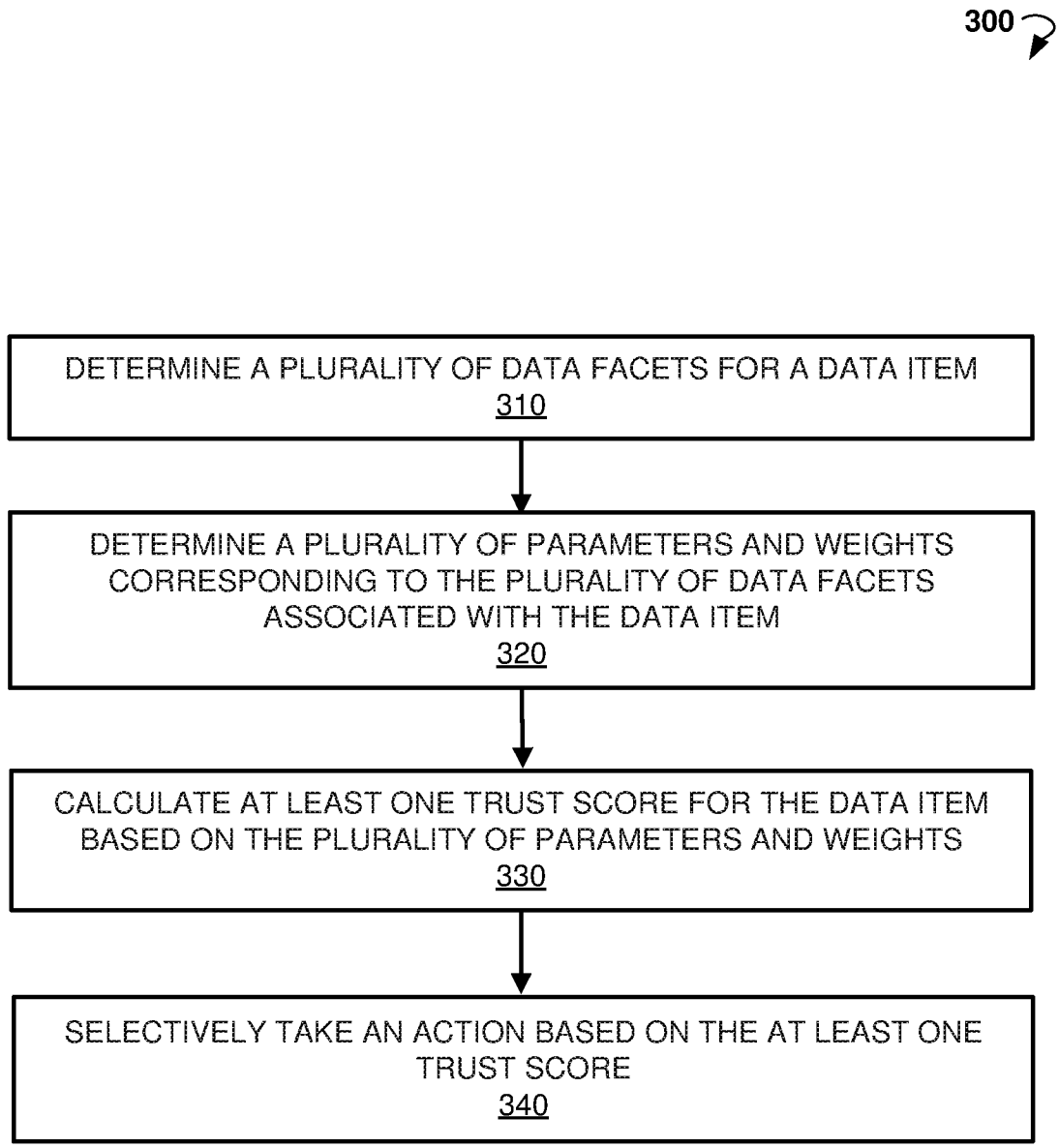
300
DETERMINE A PLURALITY OF DATA FACETS FOR A DATA ITEM
310
DETERMINE A PLURALITY OF PARAMETERS AND WEIGHTS CORRESPONDING TO THE PLURALITY OF DATA FACETS ASSOCIATED WITH THE DATA ITEM
320
CALCULATE AT LEAST ONE TRUST SCORE FOR THE DATA ITEM BASED ON THE PLURALITY OF PARAMETERS AND WEIGHTS
330
SELECTIVELY TAKE AN ACTION BASED ON THE AT LEAST ONE TRUST SCORE
340

FIG. 3

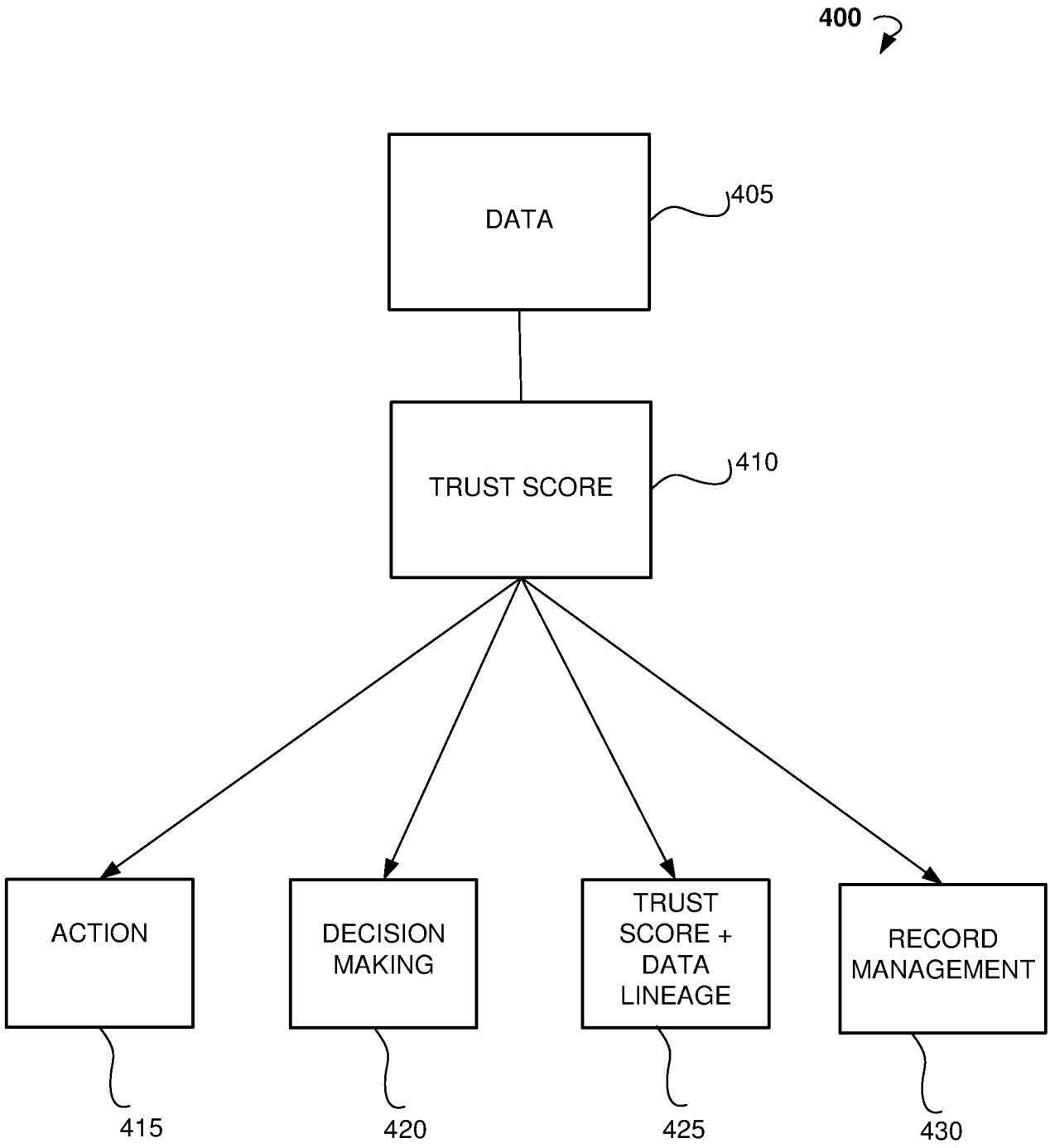
400
DATA
405
TRUST SCORE
410
ACTION
DECISION MAKING
TRUST SCORE + DATA LINEAGE
RECORD MANAGEMENT
415
420
425
430

FIG. 4

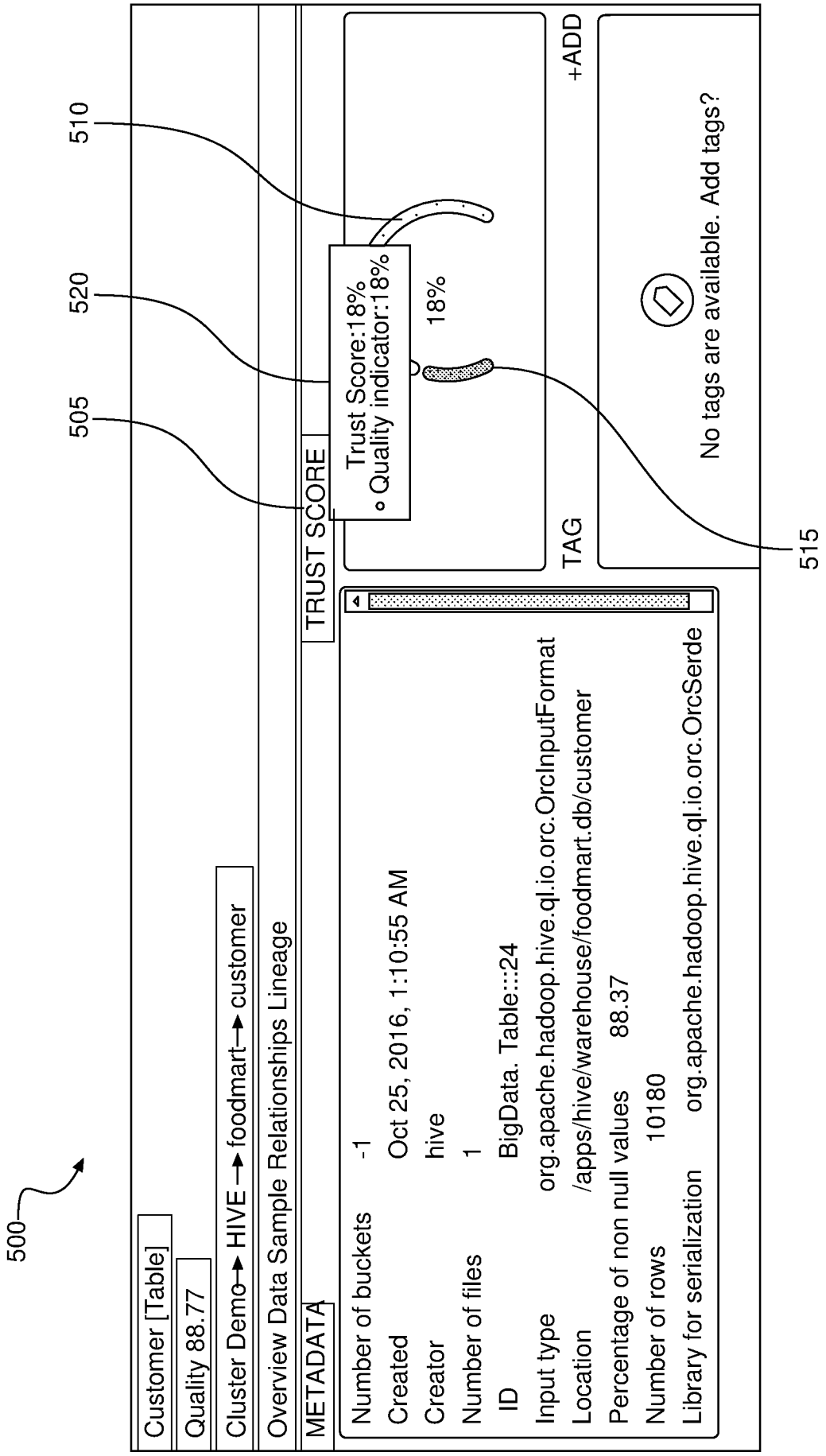
500
Customer [Table]
Quality 88.77
Cluster Demo→ HIVE→ foodmart→ customer
Overview Data Sample Relationships Lineage
METADATA
TRUST SCORE
TAG
+ADD
Trust Score:18%
Quality indicator:18%
18%
No tags are available. Add tags?
Number of buckets -1
Created Oct 25, 2016, 1:10:55 AM
Creator hive
Number of files 1
ID BigData. Table:::24
Input type org.apache.hadoop.hive.ql.io.orc.OrcInputFormat
Location /apps/hive/warehouse/foodmart.db/customer
Percentage of non null values 88.37
Number of rows 10180
Library for serialization org.apache.hadoop.hive.ql.io.orc.OrcSerde
505
510
515
520

FIG. 5

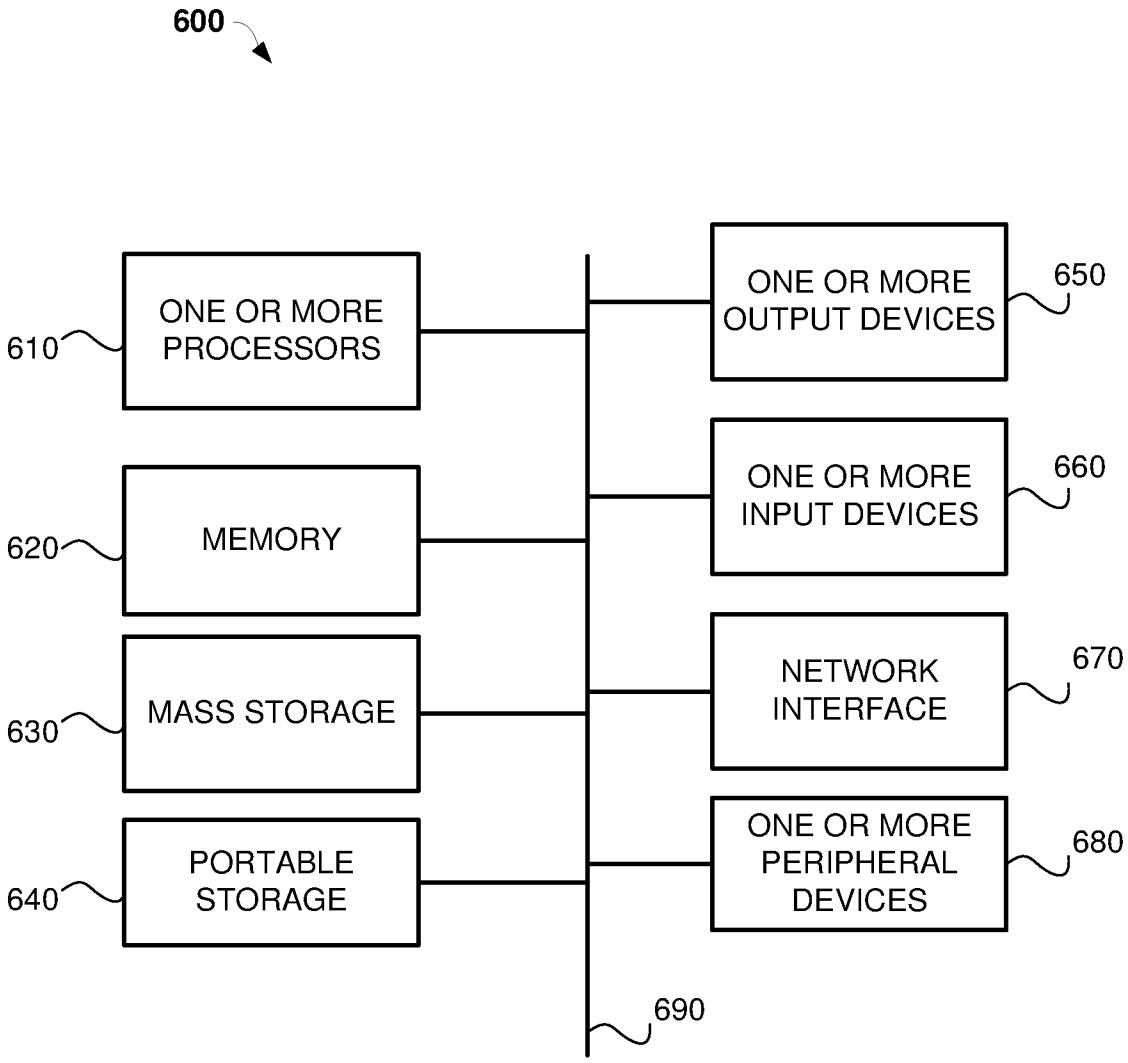
600
ONE OR MORE PROCESSORS
610
MEMORY
620
MASS STORAGE
630
PORTABLE STORAGE
640
ONE OR MORE OUTPUT DEVICES
650
ONE OR MORE INPUT DEVICES
660
NETWORK INTERFACE
670
ONE OR MORE PERIPHERAL DEVICES
680
690

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 20/56026

A. CLASSIFICATION OF SUBJECT MATTER

IPC - G06F 7/04 (2020.01)

CPC - G06F 21/10, G11B 20/00086, H04L 63/0428, H04L 2463/101, H04L 63/12, H04L 63/20, H04L 63/102, G06F 21/6218, G06F 21/604, H04L 63/0227, G06F 21/577, H04L 63/1433, H04L 63/1416, H04L 63/1408

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2010/0250497 A1 (REDLICH et al.) 30 September 2010 (30.09.2010), entire document, especially Abstract; para [0127]; [0133]; [0180]; [0187]; [0201]; [0213]; [0232]; [0238]; [0259]; [0267]; [0300]; [0318-0319]; [0334]; [0337]; [0355]; [0375]; Fig 32	1-20
A	US 2018/0191761 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 05 July 2018 (05.07.2018), entire document	1-20
A	US 2017/0357814 A1 (LOOKOUT, INC.) 14 December 2017 (14.12.2017), entire document	1-20

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 December 2020 (16.12.2020)	19 JAN 2021
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee Young Telephone No. PCT Helpdesk: 571-272-4300