



(51) International Patent Classification:

G06F 16/25 (2019.01) G06Q 10/10 (2012.01)
G06F 40/106 (2020.01)

(21) International Application Number:

PCT/MY2020/050173

(22) International Filing Date:

26 November 2020 (26.11.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

PI 2020004411 26 August 2020 (26.08.2020) MY

(71) Applicant: **MIMOS BERHAD** [MY/MY]; Technology
Park Malaysia, Kuala Lumpur, 57000 (MY).

(72) Inventors: **ABD AZIZ, Norazah**; Mimos Berhad, Technology Park Malaysia, Kuala Lumpur, 57000 (MY). **MOHD KHALID, Mohd Aminudin**; Mimos Berhad, Technology Park Malaysia, Kuala Lumpur, 57000 (MY).

(74) Agent: **MIRANDAH ASIA (MALAYSIA) SDN BHD**; Suite 3B-19-3, Plaza Sentral, Jalan Stesen Sentral 5, Kuala Lumpur, 50470 (MY).

(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,

(54) Title: A SYSTEM AND METHOD TO GENERATE STATISTICAL AND ANALYTICAL REPORT

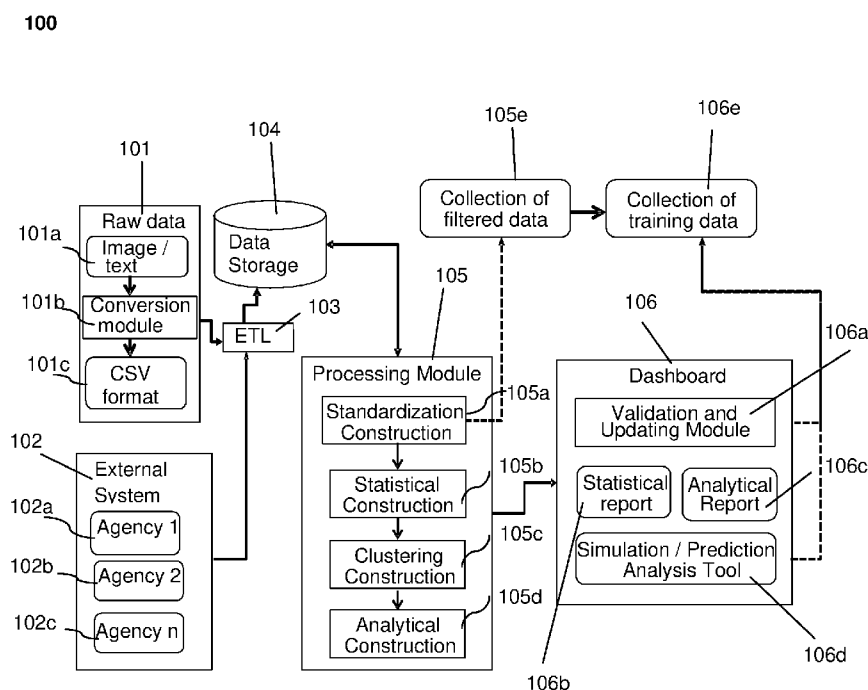


Figure 1.0

(57) Abstract: The present invention provides a system (100) and method (210) to generate a statistical and analytical report. The invention provides an construction function based on predefined attributes, classification and clustering means. The construction function involves four main modules which are a standardization construction module (105a), statistical construction module (105b), clustering construction module (105c) and an analytical construction module (105d). The standardization construction module (105a) functions to standardize input data of different formats using predefined attributes and standard deviation. The statistical construction module (105b) constructs a statistical report using classification means of predefined dataset. The clustering construction module (105c) functions as a means of classification of tagged data. The analytical construction module (105d) functions to generate an analytical report by using a clustering means and the statistical report which is then validated to produce a finalized report.



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).

Declarations under Rule 4.17:

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

Published:

- *with international search report (Art. 21(3))*
- *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

A SYSTEM AND METHOD TO GENERATE STATISTICAL AND ANALYTICAL REPORT

FIELD OF INVENTION

5

The present invention relates to a system and method to generate statistical and analytical report. In particular, the present invention provides a construction function based on predefined attributes, classification and clustering means.

10 BACKGROUND ART

Generally, most companies have their own data system and a format built in that is limited to functioning within a particular company. Hence it is not possible when data from other agencies or other companies having different formats of data to integrate into the system of a company having data of different format to produce a statistical and analytical report. The statistical and analytical report is a form of methodology provided for analysing information and planning of a company's management. In order to produce the statistical and analytical report that integrates two or more data from different agencies or companies, an expert in the field may have to predefine datasets of the two or more data from different agencies to produce the statistical and analytical report. This however, pose problems as new sets of data and requirements from each agency and company may be constantly updated into its own data system resulting in a need for an expert in the field to constantly update the predefine datasets manually.

25 The current practice to generalize a report, is either manually by an expert of the field or by using predefined dataset. In each case, a user needs to identify a type of report and data required to generate the report. A developer of the system will then have to code it accordingly. If a new data that was not previously defined is updated into the system, a normal user that is not an expert in the field will not be able to generalize the datasets and produce the statistical or analytical report. The task will therefore be considered as difficult to a normal user. Under normal circumstances, the normal user will require assistance from the developer of the system, a data storage administrator or an expert in the field to generate a new report. This method will undoubtedly be costly and time consuming.

Numerous concepts and methods have been developed to organize and analyse data, resulting in generation of analytical reports.

One example of a concept and method for generating an analytical report is disclosed in United States Patent Publication No. US 2017/0046409 A1, hereinafter referred to as US 409 A1 entitled "Using Cloud Processing to Integrate ETL into an Analytic Reporting Mechanism" having a filing date of 10 August 2015; Applicant: International Business Machines Corporation. US 409 A1 relates to a method and associated systems for using cloud processing to integrate ETL into an analytic reporting mechanism. The invention discloses a reporting tool that incorporates ETL and is used in inferring a data query as a function of receiving information input. US 409 A1 also suggest that the reporting tool is a standalone reporting tool that is used in automatically generating reports. The invention allows a reporting mechanism to create its own data-transformation requirements and use those requirements to configure and perform ETL operations that are tailored to the needs of the report. The invention further uses a cloud-computing functionality to create a distinct database for multiple request reports.

Another example of a concept and method for generating an analytical report is disclosed in United States Patent Publication No. US 2013/0238551 A1, hereinafter referred to as US 551 A1, entitled "Interest-Driven Business Intelligence Systems and Methods of Data Analysis Using Interest-Driven Data Pipelines" having a filing date of 26 April 2013, Applicant: Platfora Inc. US 551 A1 discloses a business intelligent system specifically interest-driven business intelligence systems and methods of data analysis using interest-driven data pipelines. The invention accumulates raw data in a raw data storage system and an ETL process is used to extract data from data sources. The interest-driven data pipeline then filters and/or aggregates the source data based upon a schema to create reporting data. A new data of interest if not included in the interest-driven data pipeline, will cause the interest-driven business intelligence system to rebuild the interest driven data pipeline to make the data available.

A further example of a concept and method for generating an analytical report is disclosed in United States Patent No. US 9037964 B2, hereinafter referred to as US 964 B2, entitled "Providing direct manipulation of an analytics data visualization within an analytics report" having a filing date of 12 January 2012, Applicant: International Business Machines Corporation. US 964 B2 discloses a method providing direct manipulation of

analytics data visualization within an analytics report. The invention includes a system, apparatus, computer program product and a method for providing dynamic information graphic customization to an analytics report. The event is a user interface input and the chart can be information graphic. US 964 B2 discloses that the analytics report can conform to JSON format and presented within the browser. Further, XML analytic reports can be converted to JSON format and be conveyed to the requesting entity as an enhanced report. The invention of US 964 B2 discloses a reporting engine that can be configured to permit object level analytics report deconstruction, manipulation and rendering.

There is need for further research on developing a concept that generates analytical reports and a system of standardization and classification of data for ease of integration of data, storing of data and generation of data into an analytical report. Hence, the present invention provides a system and a method to generate statistical and analytical report by providing a construction function based on predefined attributes, classification and clustering means. It enables a normal user to produce a statistical and analytical report that may have new datasets that were not previously defined and validate the datasets immediately without the need of expertise to produce a new report.

SUMMARY OF INVENTION

The present invention relates to a system and method to generate statistical and analytical report. In particular, the present invention provides a construction function based on predefined attributes, classification and clustering means.

One aspect of the present invention provides a system (100) to generate a statistical and an analytical report comprising at least one raw data module (101) for transmitting raw data into at least one extractor, transformer and loading (103) tool for extraction, at least one external module (102) having a plurality of agents for transmitting data, at least one data storage for storing data (104), at least one processing module (105) for processing the data stored in the at least one data storage and at least one validation and updating module (106a) in a dashboard (106) for validating and analysing an analytical report. The processing module (105) comprises a standardization construction module (105a) for standardizing the data based on a set of predefined attributes and standard deviation, a statistical construction module (105b) for construction of the statistical report by classification of the set of predefined attributes of the data, a clustering construction module (105c) for receiving the data and tagging the data based on a plurality of clustering reports by a clustering means and an analytical construction module (105d) for constructing an analytical report based on a combination of the statistical report and the clustering means.

Another aspect of the present invention provides a system wherein the at least one raw data module (101) comprises at least one conversion module (101b) for converting raw data in image or text format (101a) to Comma-Separated Value, CSV format (101c).

Yet another aspect of the present invention provides a method (210) for generating a statistical and an analytical report comprising receiving data having a plurality of format types (212), standardizing the data received from the plurality of format types (214), constructing the statistical report by classification means of a predefined attribute and standard deviation (216), constructing the analytical report (218,) based on the statistical report and a clustering means, validating and analysing the analytical report by using a simulation and prediction analysis tool (220). The step for standardizing data from various formats (214) comprises steps of (300) introducing raw data in CSV format or image format (302) into CSV format and further determining if raw data is in image or CSV

format (304). If raw data is in image format the image is validated, scanned and converted to CSV format (304a). If raw data is in CSV format, executing an extract, transform and load, ETL tool (306) is used for extraction of data from step (304a). Data in CSV format is extracted by executing machine readable instructions producing extracted data from an external system (306a) by using an extract, transform and load, ETL tool (306) and further storing in a first data storage (306b), executing an extract, transform and load ETL tool (306) for extraction of data from step (304a) and step (306a), processing extracted data by comparing the extracted data with at least one predefined attribute for standardization (308) and determining if extracted data can be standardized by determining if data matches any predefined attribute (310). If the data matches any predefined attribute it is updated with a predefined Identification (ID) for each of the at least one predefined attribute (316). The extracted data that have been matched with the at least one predefined attribute is checked for duplication for data storage (318) and determined if there is duplication (320). If there is no duplication the data is tagged with a profile datatype based on the predefined ID (322) and further stored in a second data storage (322b), else if there is duplication the profiling of datatype will end and will not be stored in the second data storage. If there is no match for the extracted data with the at least one predefined attribute, the extracted data is matched to values by standard deviation (312) and is determined if the data is matched in an accepted range (314). If the data matching is accepted steps (316), (318) and (320) are reiterated. If the data matching is out of accepted range, the data is stored as filtered data (314a) as a collection of filtered data for analysis (314b) and identifying a plurality of reports based on the profile datatype.

Another aspect of the present invention provides a method (210) wherein standardizing the data from various formats types (214) comprises mapping data to a set of new standard values through normalization matrix.

Yet another aspect of the present invention provides a method (210), wherein constructing the statistical report by using classification means of predefined attribute and standard deviation (216) comprises steps of (400) executing a statistical module using auto or manual interaction (402), reading the profile datatype based on the tagging of the extracted data and a classification output (404, 504), generating a plurality of reports comprising a plurality of tables and charts based on the profile datatype (406), validating the report comprising the plurality of tables and charts (408) and publishing the statistical report (410).

Another aspect of the present invention provides a method (210) wherein reading the profile datatype based on the tagging of the extracted data and classification output (404, 504), the classification output comprises steps of (500) reading the profile datatype based on the tagging of the extracted data while simultaneously checking a predefined data classification list (504), determining if a combination of the tagging of the extracted data to the predefined data classification list exist (506), classifying the extracted data based on the predefined data classification list (508) to form a new combination of the tagging of the extracted data to the predefined data classification list if the combination of the tagging of the extracted data to the predefined data classification list does not exist, storing into the second data storage for defining future extracted datasets (501), reading and selecting a clustering method for the combination of the tagging of the extracted data to the predefined data classification list (510) if the combination of the tagging of the extracted data to the predefined data classification list exist and tagging the extracted data based on the selected clustering method (512).

Yet another aspect of the present invention provides a method (210), wherein constructing the analytical report based on the statistical report and a clustering means (218,) comprises steps of (600) executing an analytical module through machine readable instructions by auto or manual interaction (602), reading and selecting the statistical report based on tagging and configuration type by forming and rearranging the statistical report (604), reading the data based on a plurality of clustering type of reports (606), generating analytical datasets for the analytical report , wherein the analytical report covers accuracy, precision, linearity and specificity (608), validating and modifying the analytical reports (610), determining if the analytical reports are accepted (612) and publishing the analytical reports (614), determining if the analytical reports are not accepted (612), storing the analytical report as training data for further analysis (616), and collecting the analytical report as a collection of training data (618).

Another aspect of the present invention provides a method (210) wherein generating the analytical datasets for the analytical report (608) comprises generating an analytical report, tables and charts.

The present invention consists of features and a combination of parts hereinafter fully described and illustrated in the accompanying drawings, it being understood that various

changes in the details may be made without departing from the scope of the invention or sacrificing any of the advantages of the present invention.

BRIEF DESCRIPTION OF ACCOMPANYING DRAWINGS

To further clarify various aspects of some embodiments of the present invention, a more particular description of the invention will be rendered by references to specific
5 embodiments thereof, which are illustrated in the appended drawings. It is appreciated that these drawings depict only typical embodiments of the invention and are therefore not to be considered limiting of its scope. The invention will be described and explained with additional specificity and detail through the accompanying drawings in which:

10 Figure 1.0 illustrates a general architecture of the system having modules and sub modules for generating a statistical and analytical report.

Figure 2.0a is a schematic diagram illustrating a general overview of the general methodology of generating a statistical and analytical report of the present invention.

15

Figure 2.0b is a flowchart illustrating a general methodology for generating a statistical and analytical report of the present invention.

20 Figure 3.0 is a flowchart illustrating a method of standardization of data for generating a statistical and analytical report.

Figure 4.0 is a flowchart illustrating a method for statistical construction for generating a statistical report.

25 Figure 5.0 is a flowchart illustrating a method for clustering construction for generating a statistical and analytical report.

Figure 6.0 is a flowchart illustrates a method for analytical construction for generating an analytical report.

DETAILED DESCRIPTION OF THE DRAWINGS

The present invention relates to a system and method to generate statistical and analytical report. In particular, the present invention provides a construction function based on predefined attributes, classification and clustering means. Hereinafter, this specification will describe the present invention according to the preferred embodiments. It is to be understood that limiting the description to the preferred embodiments of the invention is merely to facilitate discussion of the present invention and it is envisioned without departing from the scope of the appended claims.

Reference is first made to Figure 1.0. Figure 1.0 illustrates a general architecture of a system having modules and sub modules for generating a statistical and analytical report (100). The system (100) comprises at least one raw data module (101) for transmitting raw data into at least one extractor, transformer and loading (ETL) (103) tool for extraction, at least one external module (102) having a plurality of agents for transmitting data, at least one data storage for storing data (104), at least one processing module (105) for processing the data stored in the at least one data storage, at least one validation and updating module (106a) in a dashboard (106) for validating and analysing an analytical report. The data is selected from more than one type of formats. For example the formats where data is selected may be in Comma-Separated Value, CSV format or image format. The processing module (105) comprises a standardization construction module (105a) for standardizing the data based on a set of predefined attributes and standard deviation, a statistical construction module (105b) for construction of the statistical report by classification of the set of predefined attributes of the data, a clustering construction module (105c) for receiving the data and tagging the data based on a plurality of clustering reports by a clustering means and an analytical construction module (105d) for constructing an analytical report based on a combination of the statistical report and the clustering means. The at least one raw data module comprises at least one conversion module (101b) for converting data in image or text format (101a) to CSV format (101c). The ETL functions to extract data from external systems (102) based on a trigger procedure configuration. Any new incoming data triggers the computing and initializing of a standardization construction function. The ETL processes uses ETL tools to extract data. These ETL tools may be Pentaho, Jasper, GeoKettle but not limited to these ETL tools.

The plurality of sub modules of the processing module (105) comprises a standardization construction module (105a) for standardizing the data based on a set of predefined attributes and standard deviation, a statistical construction module (105b) for construction of a statistical report by classifying the set of predefined attributes of the data, a clustering construction module (105c) for receiving the data and tagging the data based on a plurality of clustering reports by a clustering means and an analytical construction module (105d) for constructing an analytical report based on a combination of the statistical report and the clustering means.

10 The system is initialized when data is received from various sources including raw data and processed data. The raw data (101) may be in text or image format (101a) which is then processed by the conversion module (101b) to a standardized CSV format (101c). By using an ETL tool (103) the standardized data in CSV format is extracted and stored in the data storage (104). The ETL also extracts data from the external system module (102) consisting of various agencies such as Agency 1 (102a), Agency 2 (102b) and Agency n (102c) that have other data storages that manages and processes separately from the system of the present invention. The external system (102) further stores data in the data storage (104) of the present invention to be used in generating a report.

20 The standardization construction module (105a) processes incoming data by comparing predefined attributes for standardization. The data is matched to a predefined attribute and is updated. If the data does not match with a predefined attribute, the data will be matched to a possible predefined attribute value and updated. The standardization data with the predefined attribute is then tagged to a profile of the datatype also known as a tagging attribute. Some of the data is collected and stored as filtered data (105e) or training data (106e).

The statistical construction module (105b) is used to generate statistical report based on the tagging attribute and classification output. Tables and charts are generated based on the tagged profile and is verified by a verification tool before publishing the statistical report.

The clustering construction module (105c) is used as a means of classification of datatype. The tagged data is read by the system and checks the predefined data classification list by checking the tagging combination with the predefined data classification. If the tagging

combination with the classification exists, possible clustering groupings are determined. If the tagging combination with the classification does not exist, new combinations of the tagging and new classification are formed and the data is tagged on possible clustering groupings.

5

The analytical construction module (105d) serves to create an analytical report by reading the data based on classification tagging involving the selection of statistical report and reading the data to determine possible clustering type of report in order to construct analytical reports comprising tables and charts.

10

The last module is the validation and updating module (106a) to generate the final report. The statistical report (106b) produced is used to produce the analytical report (106c) that is analysed and validated. A simulation and prediction analysis tool (106d) is used as an aid in producing a good analytical report. Once validated by the validation module (106a) the analytical report is published. If validation does not initialize, the dataset is stored as training data (106e) for further analysis.

15

Reference is made to Figure 2.0a. Figure 2.0a illustrates a general overview of a method of generating a statistical and analytical report (200). There are two data inputs to initiate the system in generating the statistical and analytical report. One type of data is data of different formats (201) extracted from data storages of different bodies or agencies. These set of data are either raw or processed data having different formats. Another type of data is existing ones from a data storage of a body or agency (202). For data having different formats a module of collection and standardization of data (203) will be initiated to standardize the data before grouping the data by classification and clustering means (204). The data is then processed to produce the statistical report (205). Finally, an analytical report is generated (206) based on the statistical report and clustering means.

20

25

In describing further on the method of generating a statistical and analytical report reference is also made to Figure 2.0b. Figure 2.0b discloses a method (210) for generating a statistical and an analytical report. The method comprises receiving data having a plurality of format types (212), standardizing the data received from the plurality of format types (214), constructing the statistical report by classification means of a predefined attribute and standard deviation (216), constructing the analytical report (218)

30

based on the statistical report and a clustering means and validating and analysing the analytical report by using a simulation and prediction analysis tool (220).

Reference is made to Figure 3.0. Figure 3.0 illustrates a method of standardization of data for generating a statistical and analytical report. Standardizing the data received from the plurality of format types comprises (300) introducing at least one raw data in CSV format or image format for converting raw or processed data to CSV format (302) and determining if the raw data is in image or CSV format (304). If the raw data is in image format the image is validated, scanned and converted to CSV format (304a). If raw data is in CSV format, an extract, transform and load, ETL tool is executed for extraction of data from step (304a). Data in CSV format is extracted by executing machine readable instructions producing extracted data from an external system (306a) by using an extract, transform and load, ETL tool (306) and further storing in a first data storage (306b). The extract, transform and load ETL tool (306) is executed for extraction of data from step (304a) and step (306a) and the extracted data is processed by comparing the extracted data with at least one predefined attribute for standardization (308) and determining if extracted data can be standardized by determining if data matches any predefined attribute (310). If the data matches any predefined attribute it is updated with a predefined ID for each of the at least one predefined attribute (316). The extracted data that have been matched is checked with the at least one predefined attribute for duplication for data storage (318) and further determined if there is any duplication in matching of the extracted data with the at least one predefined attribute (320). The extracted data that have no duplication is tagged with a profile datatype based on the predefined ID (322) and the tagged extracted data is stored into second data storage (322b). If there is no match for the extracted data with the at least one predefined attribute, the extracted data is matched to values by standard deviation (312).

In initializing a data collection module comprising a conversion module for converting raw or processed data to CSV format (302), it involves determining if the data is in image or CSV format (304). If the format is in image format, scanning and processing of the image is initialized to convert the image into CSV format (304a). Further in the extraction of data that initializes at least one machine readable instruction that run continuously to extract data in CSV format from an external system (306a) by using an extract, transform and load, ETL tool (306), the extracted data is also stored in a first data storage (306b) that runs a data storage query to compare predefined attributes. In an event where a negative

result is generated when determining if a data can be standardized (310), the standard deviation program code or script is run to compare with the predefined attribute listed for possible data where the extracted data is matched to values by standard deviation (312). By using a normalization matrix, the data will be identified to the possible data matched as it maps dirty data to new standard data values. Thereafter, it is determined if the data matched is in an accepted range (314). If the data matches, data is updated with a predefined ID for each of the at least one predefined attribute (316). The extracted data that have been matched with the at least one predefined attribute for duplication for data storage is checked for duplication (318), The updated standardization data is checked further for duplication (320) and is tagged with a profile datatype based on the predefined ID (322). The data is then stored in a second data storage (322b). If there is duplication (320) the profiling of datatype will end and will not be stored in the second data storage to avoid redundancy and to eliminate duplication of data. If the data matching is out of the accepted range (314), the data that does not match is stored as filtered data (314a) as a collection of filtered data for analysis (314b).

Some examples of attributes of the present invention include year, month, gender, state name and race but are not limited to the said examples of attributes alone. Further, tagging of a profile datatype based on predefined ID includes report categories that may consist of examples like population, economy, education, assistance but not limited to the said examples of predefined IDs.

Reference is now made to Figure 4.0. Figure 4.0 illustrates a method for statistical construction for generating a statistical report. Constructing the statistical report using classification means of predefined attribute and standard deviation (400) comprises steps of executing a statistical construction module by executing the statistical module by auto or manual interaction (402), reading the profile datatype based on the tagging of the extracted data and a classification output (404), also referenced as (404a), generating a plurality of reports comprising a plurality of tables and charts based on the profile datatype (406), validating the report comprising the plurality of tables and charts (408), and publishing the statistical report (410). The second data storage (402a) comprises previously stored data that was extracted by the ETL tool.

The classification means of the present invention involves a method of learning a model that categorizes different predetermined classes of data. It involves a two-step process

comprising a learning step and a classification step. The learning step can be accomplished by using an already defined training set of data. Some algorithms for classification of the data for the present invention may be Logistic Regression, Decision tree, K-nearest neighbours, KNN, Support Vector Machines, SVM and Random Forest.

5 The algorithms may is not limited to the said algorithms alone and may include any algorithms for classification that is able to categorize data for an analytical report.

Reference is now made to Figure 5.0 illustrating a method for clustering construction for generating a statistical and analytical report (500). The method involves reading the
10 profile datatype based on the tagging of the extracted data and classification output. The classification output comprises reading the profile datatype (502) referenced earlier as (404) and (404a) and reading based on the tagging of the extracted data while simultaneously checking a predefined data classification list (504), if a combination of the tagging of the extracted data to the predefined data classification list is exist as in step
15 (506), reading and selecting a clustering method for the combination of the tagging of the extracted data to the predefined data classification list (510) and tagging the extracted data based on the selected clustering method (512) before proceeding with reading of the clustering types of report (514) which will be further illustrated in Figure 6.0.

20 In an event where the combination of the tagging of the extracted data to the predefined data classification list does not exist (506), classifying the extracted data based on the predefined data classification list (508) to form a new combination of the tagging of the extracted data to the predefined data classification list and storing into a second data storage for defining future extracted datasets (501).

25 Some of the clustering methods (510b) used in the present invention include partitioning methods, hierarchical clustering, fuzzy clustering, density-based clustering and model-based clustering.

30 Reference is made to Figure 6.0 illustrating a method for analytical construction for generating an analytical report. The method of constructing an analytical report based on the statistical report and a clustering means (600) comprises steps of executing an analytical construction module through machine readable instructions by auto or manual interaction (602), reading and selecting the statistical report based on tagging and
35 configuration type by forming and rearranging the statistical report (604), reading the data

based on a plurality of clustering type of reports (606), generating analytical datasets for the analytical report, the analytical report covers accuracy, precision, linearity and specificity (608), and validating and modifying the analytical reports (610). For both the steps of reading and selecting the statistical report based on tagging and configuration type by forming and rearranging the statistical report (604) and reading the data based on a plurality of clustering type of reports (606), the steps are linked to a second data storage (601) having an analytical dataset library collection (601a). The analytical report (608) comprises step of generating an analytical report, tables and charts. In validating and modifying the analytical reports (610), validation of the reports comprises steps of validating the analytical reports by identifying accepting of the analytical reports (612), and publishing the analytical reports (614). Some of the tools (610a) used in validating and modifying the reports, tables or charts may be decision maker tools, simulation tools and analytic tools but may not be limited to these validation tools alone and may include other validation tools.

In an event where validating and modifying the analytical reports (610) comprises not accepting the analytical reports (612), the flow of analytical construction proceeds with storing the analytical reports as training data for further analysis (616) and collecting the analytical reports as a collection of training data (618).

Throughout this specification, unless the context requires otherwise, the word “comprise”, or variations such as “comprises” or “comprising”, will be understood to imply the inclusion of a stated step or element or integer or group of steps or elements or integers, but not the exclusion of any other step or element or integer or group of steps, elements or integers. Thus, in the context of this specification, the term “comprising” is used in an inclusive sense and thus should be understood as meaning “including principally, but not necessarily solely”.

CLAIMS

1. A system (100) to generate a statistical and an analytical report comprising:
- at least one raw data module (101) for transmitting raw data into at least one extractor, transformer and loading (103) tool for extraction;
 - 5 at least one external module (102) having a plurality of agents for transmitting data;
 - at least one data storage for storing data (104);
 - at least one processing module (105) for processing the data stored in the at least one data storage (104); and
 - 10 at least one validation and updating module (106a) in a dashboard (106) for validating and analysing the analytical report,
- characterized in that
- the processing module (105) comprises:
 - 15 a standardization construction module (105a) for standardizing the data based on a set of predefined attributes and standard deviation;
 - a statistical construction module (105b) for construction of the statistical report by classification of the set of predefined attributes of the data;
 - 20 a clustering construction module (105c) for receiving the data and tagging data based on a plurality of clustering reports by clustering means; and
 - an analytical construction module (105d) for constructing the analytical report based on a combination of the statistical report and the clustering means.
 - 25
2. The system according to claim 1, wherein the at least one raw data module (101) comprises at least one conversion module (101b) for converting raw data in image or text format (101a) to Comma-Separated Value, CSV format (101c).
- 30
3. A method (210) for generating a statistical and an analytical report comprising:
- receiving data having a plurality of format types (212);
 - standardizing the data received from the plurality of format types (214);
 - constructing the statistical report by classification means of a predefined attribute and standard deviation (216);
 - 35

constructing the analytical report (218) based on the statistical report and a clustering means; and

validating and analysing the analytical report by using a simulation and prediction analysis tool (220),

5 characterized in that

standardizing the data received from the plurality of format types (214) comprises steps of (300):

introducing raw data in CSV format or image format (302);

determining if raw data is in image or CSV format (304),

10 if raw data is in image format, validating, scanning and converting the image format to CSV format (304a);

else, executing an extract, transform and load, ETL tool (306) for extraction of data from step (304a);

15 extracting data in CSV format by executing machine readable instructions for producing extracted data from an external system (306a) by using an extract, transform and load, ETL tool (306) and storing extracted data in a first data storage (306b);

20 executing the extract, transform and load, ETL tool (306) for extraction of data from step (304a) and step (306a);

processing the extracted data by comparing the extracted data with at least one predefined attribute for standardization (308); and

25 determining if the extracted data can be standardized by determining if data matches any predefined attribute (310); wherein,

if the data matches any predefined attribute,

30 updating data with a predefined identification, ID for each of the at least one predefined attribute (316);

checking the extracted data that have been matched with the at least one predefined attribute for duplication for data storage (318);

determining if there is duplication (320);
wherein,
if there is no duplication,
tagging data with a profile datatype
based on the predefined ID (322) and
stored in second data storage (322b);
else if there is duplication,
profiling of datatype will end and will
not be stored in the second data
storage;
else if there is no match for the extracted data with
the at least one predefined attribute,
matching the extracted data to values by
standard deviation (312); and
determining if data is matched (314) in
accepted range;
wherein,
if data matching is in the accepted
range;
reiterating steps (316), (318) and
(320);
else,
storing filtered data (314a) as a
collection of filtered data for
analysis(314b).

4. The method (210) according to claim 3, wherein standardizing the data from various format types (214) comprises mapping data to a set of new standard values through normalization matrix.

5. The method (210) according to claim 3, wherein constructing the statistical report by using classification means of predefined attribute and standard deviation (216) comprises steps of (400):

executing a statistical module using auto or manual interaction (402);

reading the profile datatype based on the tagging of the extracted data and a classification output (404, 504);

generating a plurality of reports comprising a plurality of tables and charts based on the profile datatype (406);

5 validating the report comprising the plurality of tables and charts (408); and
publishing the statistical report (410).

6. The method (210) according to claim 5, wherein reading the profile datatype based on the tagging of the extracted data and classification output (404, 504), the
10 classification output comprises steps of (500):

reading the profile datatype based on the tagging of the extracted data while simultaneously checking a predefined data classification list (504);

determining if a combination of the tagging of the extracted data to the predefined data classification list exist (506);

15 classifying the extracted data based on the predefined data classification list (508) to form a new combination of the tagging of the extracted data to the predefined data classification list if the combination of the tagging of the extracted data to the predefined data classification list does not exist;

storing into the second data storage for defining future extracted datasets (501);

20 reading and selecting a clustering method for the combination of the tagging of the extracted data to the predefined data classification list (510) if the combination of the tagging of the extracted data to the predefined data classification list exist; and
tagging the extracted data based on the selected clustering method (512).

25 7. The method (210) according to claim 3, wherein constructing the analytical report based on the statistical report and a clustering means (218) comprises steps of (600):

executing an analytical module through machine readable instructions by auto or manual interaction (602);

30 reading and selecting the statistical report based on tagging and configuration type by forming and rearranging the statistical report (604);

reading the data based on a plurality of clustering type of reports (606);

generating analytical datasets for the analytical report, wherein the analytical report covers accuracy, precision, linearity and specificity (608);

validating and modifying the analytical reports (610);

determining if the analytical reports are accepted (612) and publishing the analytical reports (614);

determining if the analytical reports are not accepted (612), storing the analytical reports as training data for analysis (616); and

5 collecting the analytical reports as a collection of training data (618).

8. The method (210) of claim 7, wherein generating the analytical datasets for the analytical reports (608) comprises generating tables and charts.

100

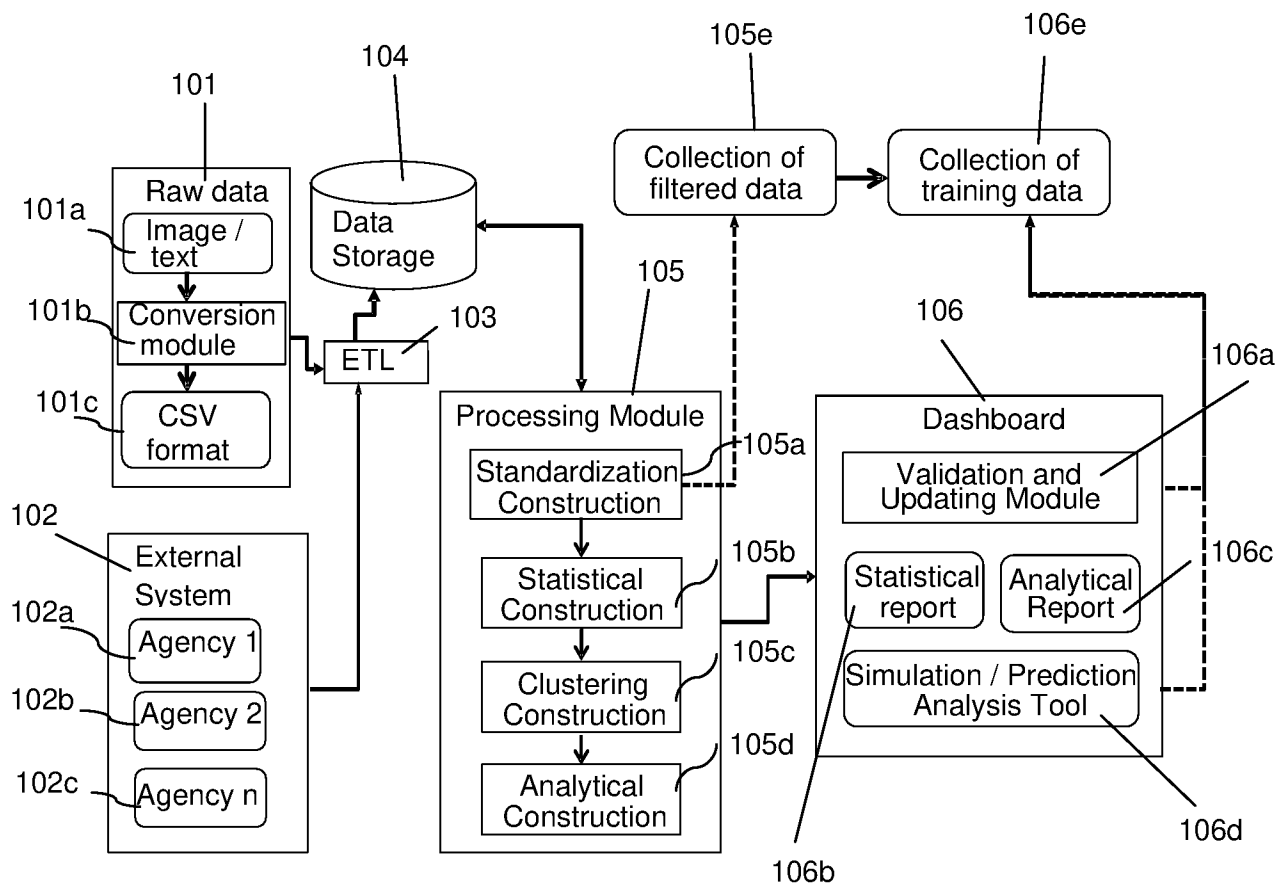


Figure 1.0

200

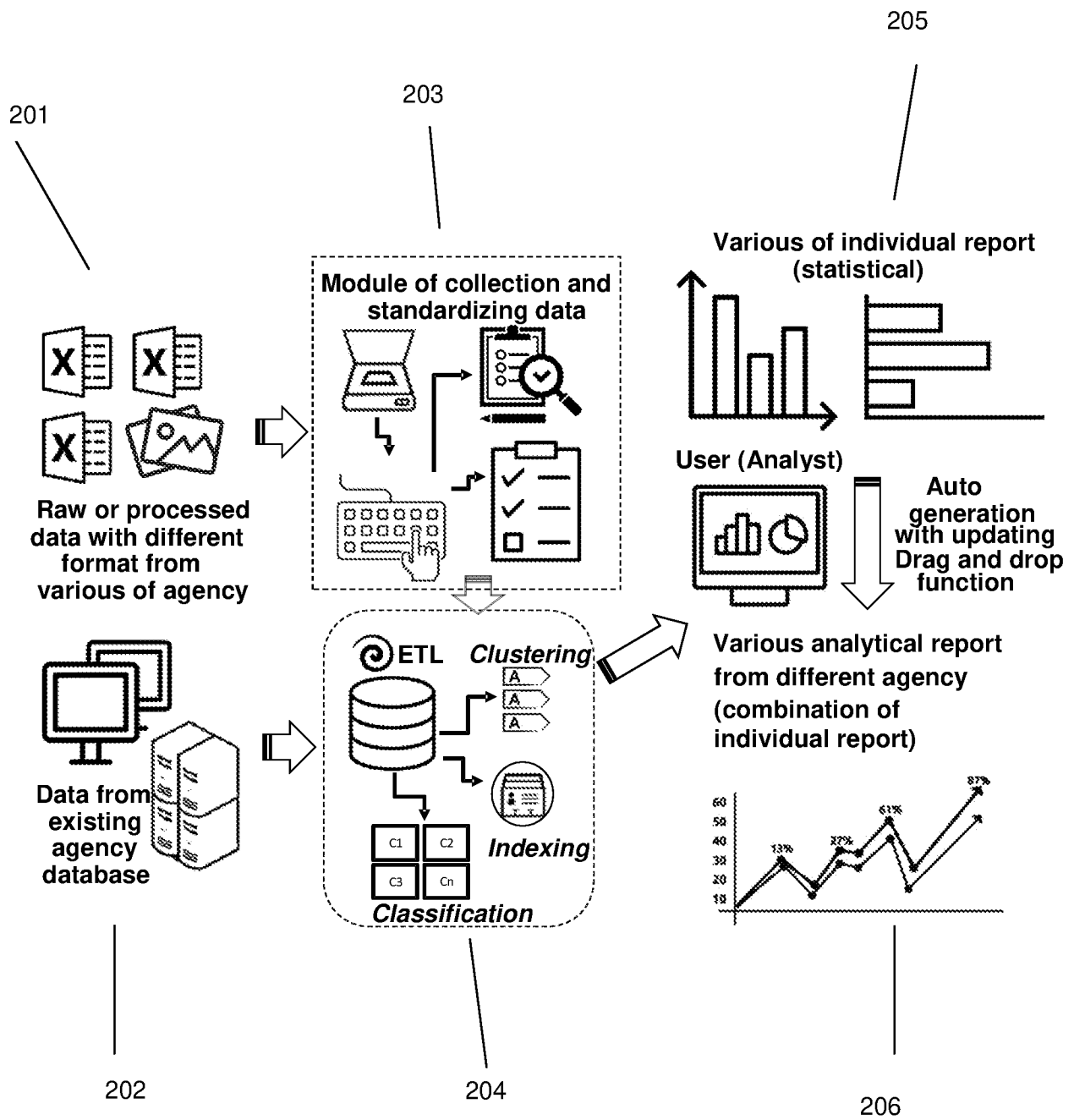


Figure 2.0a

210

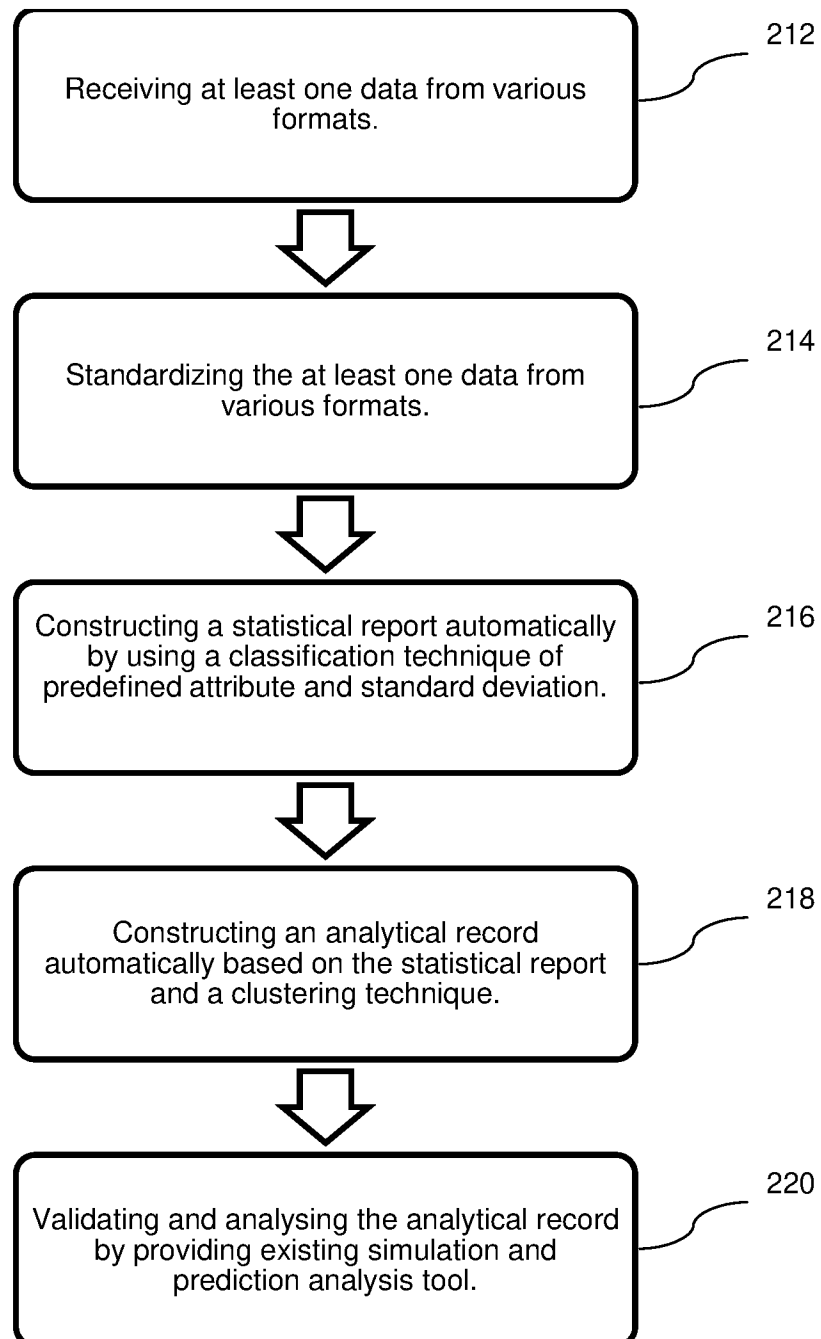


Figure 2.0b

300

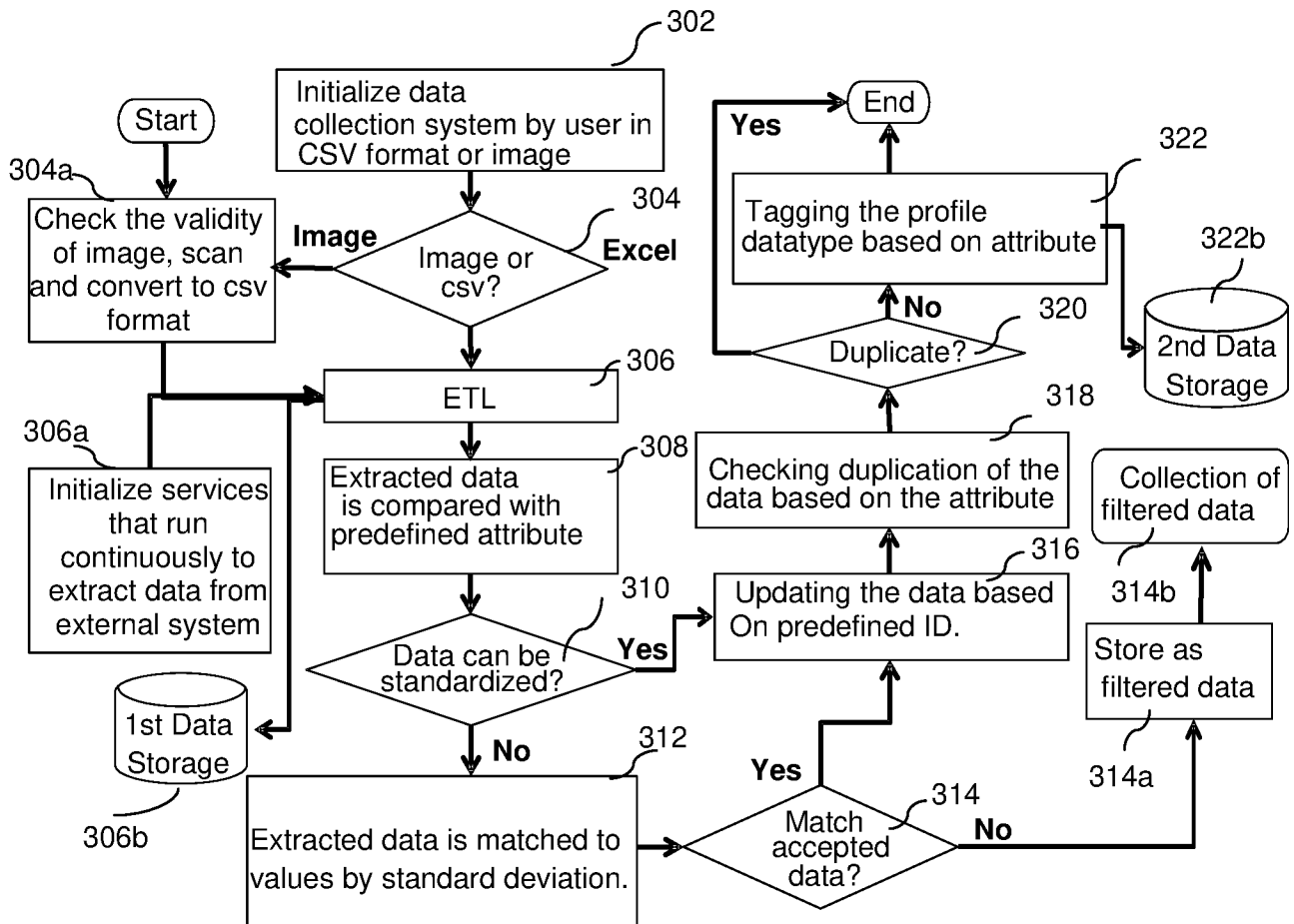


Figure 3.0

400

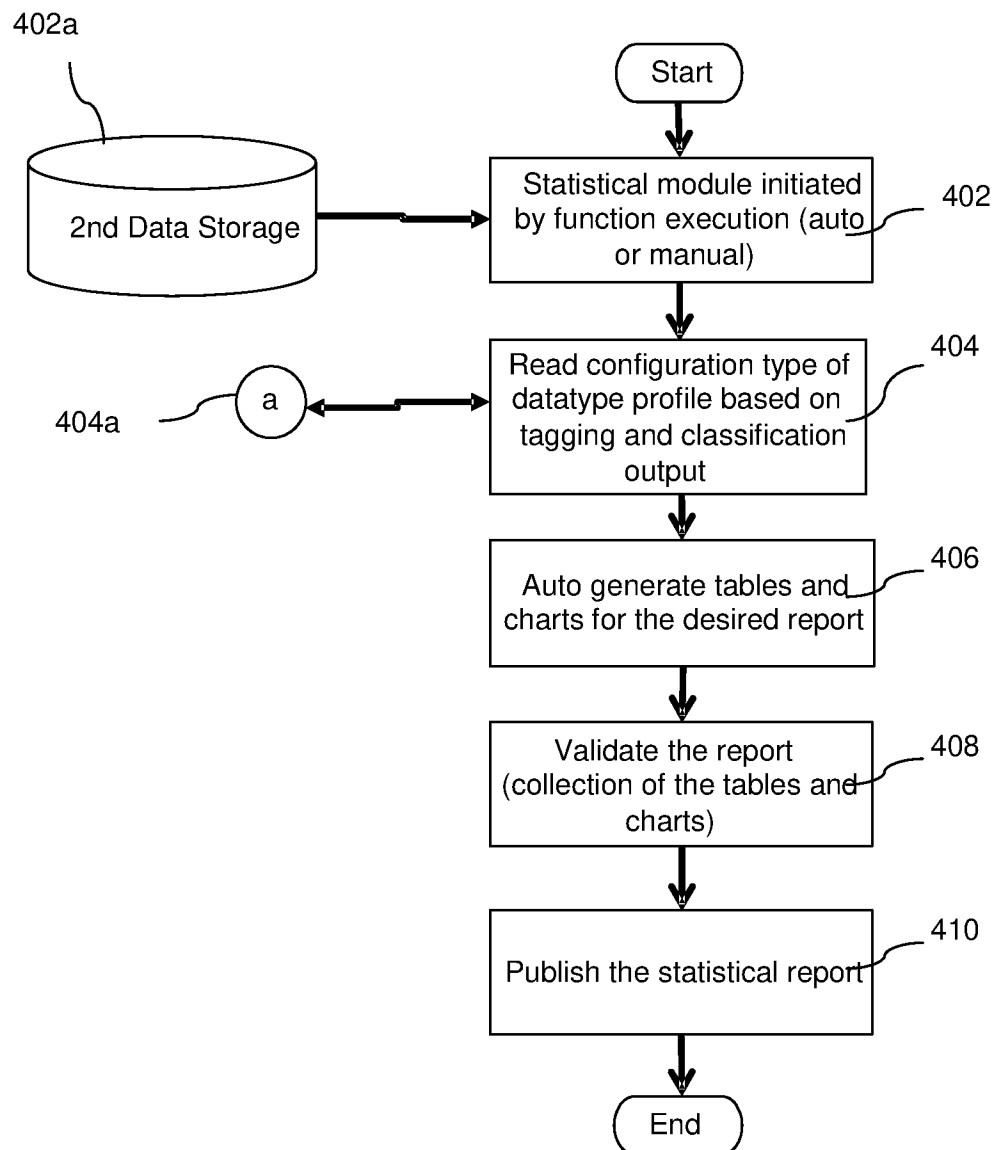


Figure 4.0

500

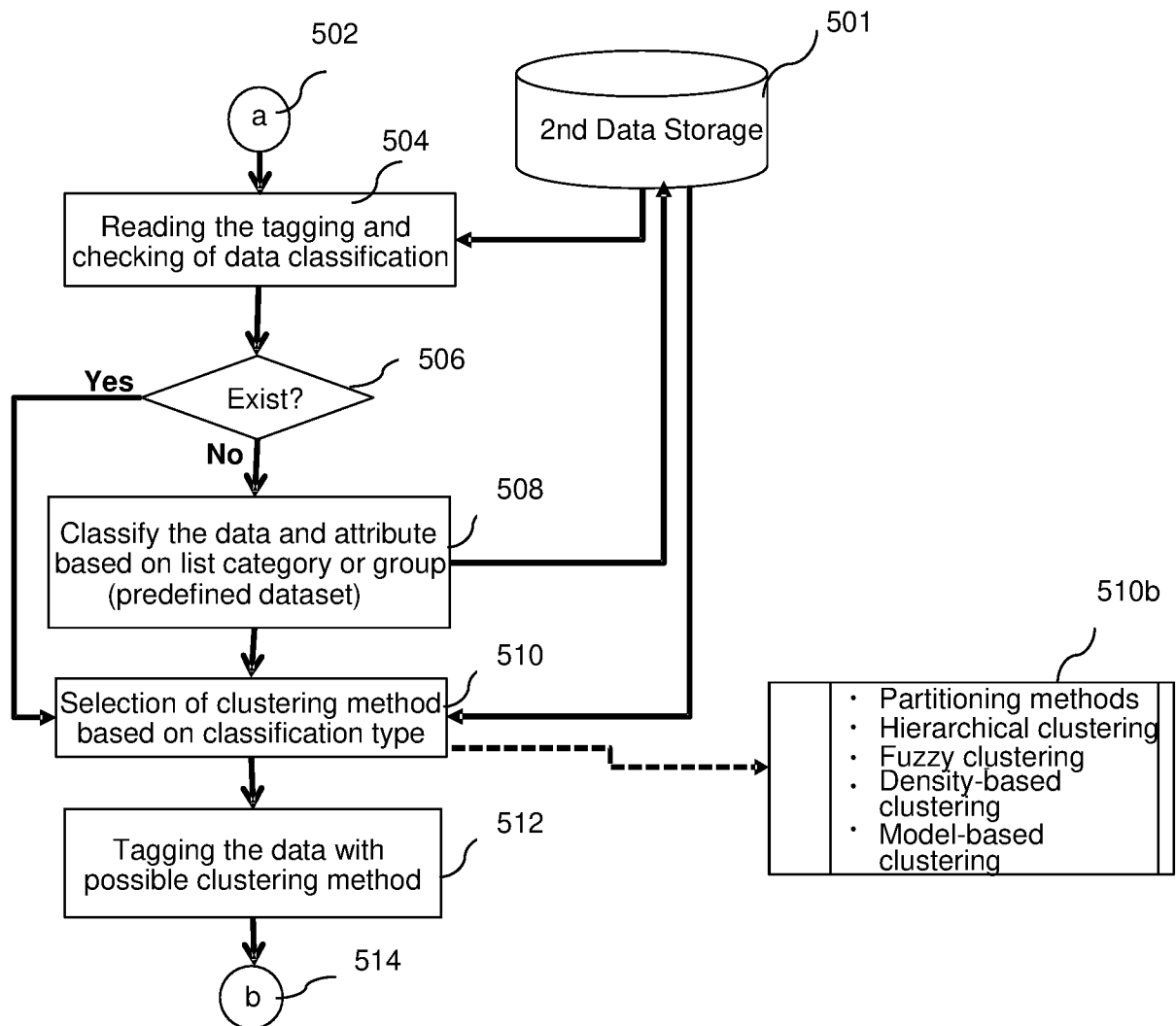


Figure 5.0

600

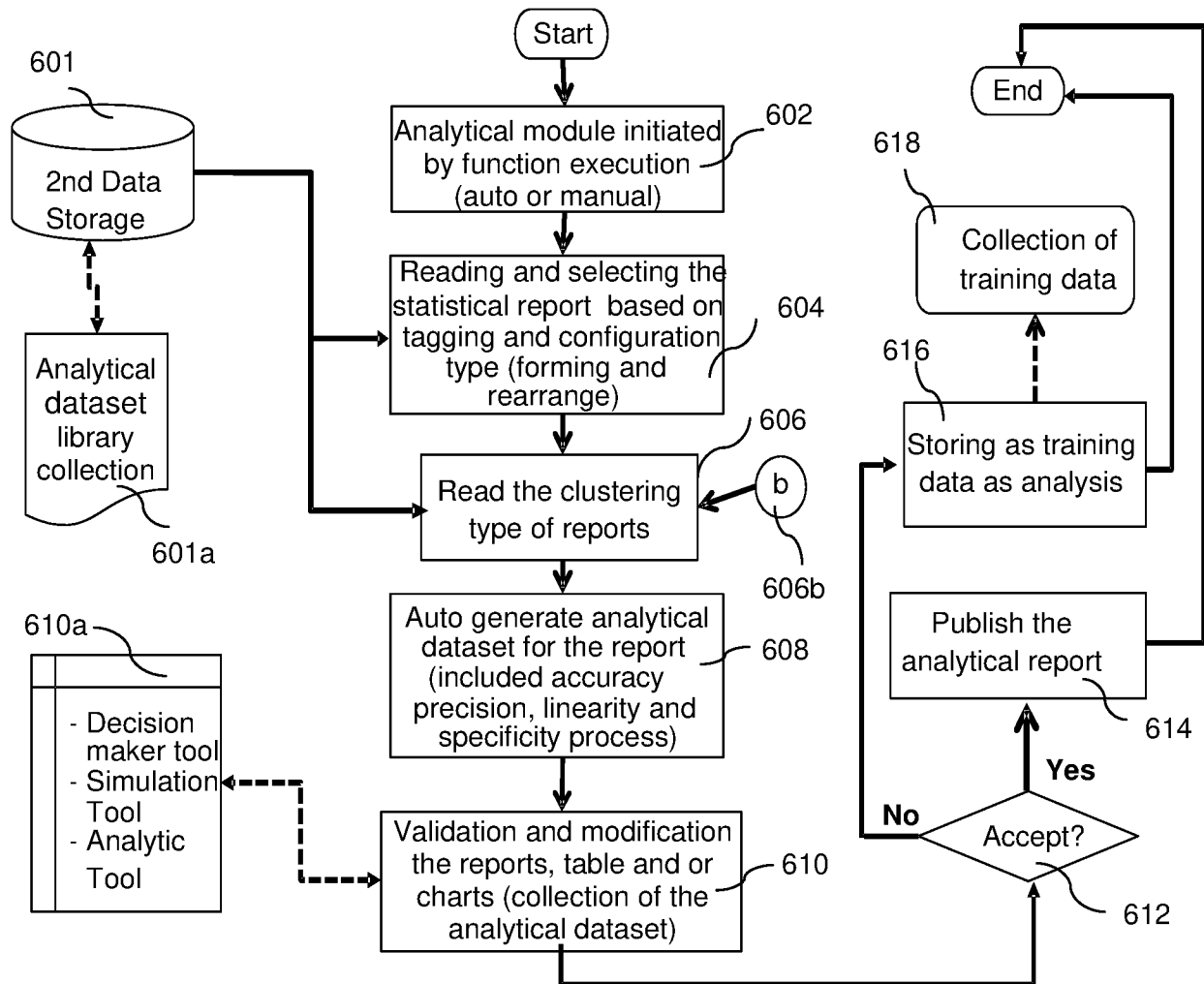


Figure 6.0

INTERNATIONAL SEARCH REPORT

International application No.

PCT/MY2020/050173

A. CLASSIFICATION OF SUBJECT MATTER G06F 16/25(2019.01)i; G06F 40/106(2020.01)i; G06Q 10/10(2012.01)i 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) G06F 16/25(2019.01); G06N 3/02(2006.01); G06Q 10/00(2006.01); G06Q 10/04(2012.01); G06Q 10/06(2012.01); G06Q 50/18(2012.01) Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Korean utility models and applications for utility models Japanese utility models and applications for utility models Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) eKOMPASS(KIPO internal) & Keywords: analysing, statistical report, clustering, standardization, external data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	US 2014-0025442 A1 (QUID, INC.) 23 January 2014 (2014-01-23) paragraphs [0027], [0030], [0034], [0035], [0049] and claim 6	1,2 3-8
Y	US 8935198 B1 (G. MICHAEL PHILLIPS et al.) 13 January 2015 (2015-01-13) claim 1	1,2
Y	US 2005-0131928 A1 (PHILIPPE GAUTHEY et al.) 16 June 2005 (2005-06-16) claim 19	2
A	US 8504408 B2 (MANISHA BANTHIA et al.) 06 August 2013 (2013-08-06) claims 1-14	1-8
A	KR 10-1443028 B1 (KOREA ATOMIC ENERGY RESEARCH INSTITUTE) 19 September 2014 (2014-09-19) paragraphs [0018]-[0123]	1-8
☐ 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 24 May 2021		Date of mailing of the international search report 24 May 2021
Name and mailing address of the ISA/KR Korean Intellectual Property Office 189 Cheongsa-ro, Seo-gu, Daejeon 35208, Republic of Korea Facsimile No. +82-42-481-8578		Authorized officer KIM, Sung Hoon Telephone No. +82-42-481-8710

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/MY2020/050173

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2014-0025442	A1	23 January 2014	BR	PI0917246	A2	10 November 2015
				CN	102150129	A	10 August 2011
				EP	2329357	A1	08 June 2011
				JP	2011-530138	A	15 December 2011
				KR	10-2011-0056502	A	30 May 2011
				US	2010-0030722	A1	04 February 2010
				US	8554709	B2	08 October 2013
				WO	2010-017229	A1	11 February 2010
US	8935198	B1	13 January 2015	US	7072863	B1	04 July 2006
				US	7958204	B1	07 June 2011
				US	8521663	B1	27 August 2013
				US	8954361	B1	10 February 2015
US	2005-0131928	A1	16 June 2005	EP	1709553	A1	11 October 2006
				US	7925662	B2	12 April 2011
				WO	2005-008542	A1	27 January 2005
US	8504408	B2	06 August 2013	US	2011-0251874	A1	13 October 2011
KR	10-1443028	B1	19 September 2014	JP	2014-010837	A	20 January 2014
				JP	5841102	B2	13 January 2016
				KR	10-2014-0001620	A	07 January 2014
				US	2014-0006294	A1	02 January 2014