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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
08 April 2021 (08.04.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/067447 A1

(51) International Patent Classification:
G06F 40/289 (2020.01) *G06F 40/56* (2020.01)
G06F 40/35 (2020.01) *G10L 13/04* (2013.01)

(21) International Application Number:
PCT/US2020/053560

(22) International Filing Date:
30 September 2020 (30.09.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/908,235 30 September 2019 (30.09.2019) US

(71) Applicant: **STATS LLC** [US/US]; 203 North Lasalle St. Suite 2200, Chicago, IL 60601 (US).

(72) Inventors: **LONG, Adam**; C/o Stats Llc, 203 North Lasalle St. Suite 2200, Chicago, IL 60601 (US). **HAYNES, Nick**; C/o Stats Llc, 203 North Lasalle St. Suite 2200, Chicago, IL 60601 (US). **SALO, Dan**; C/o Stats Llc, 203 North Lasalle St. Suite 2200, Chicago, IL 60601 (US). **DAHM, Heather**; C/o Stats Llc, 203 North Lasalle St. Suite 2200, Chicago, IL 60601 (US). **MCCRIGHT, Jacob**; C/o Stats Llc, 203 North Lasalle St. Suite 2200, Chicago, IL 60601 (US). **DILLON, Michael**; C/o Stats Llc, 203 North Lasalle St. Suite 2200, Chicago, IL 60601 (US). **DETERS, Hannah**; C/o Stats Llc, 203 North Lasalle St. Suite 2200, Chicago, IL 60601 (US). **KHADKA, Anup**; C/o Stats Llc, 203 North Lasalle St. Suite 2200, Chicago, IL 60601 (US).

(74) Agent: **WOLFE, Joseph** et al.; Dla Piper Llp Us, 33 Arch Street, 26th Floor, Boston, MA 02110 (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,

(54) Title: AUGMENTED NATURAL LANGUAGE GENERATION PLATFORM

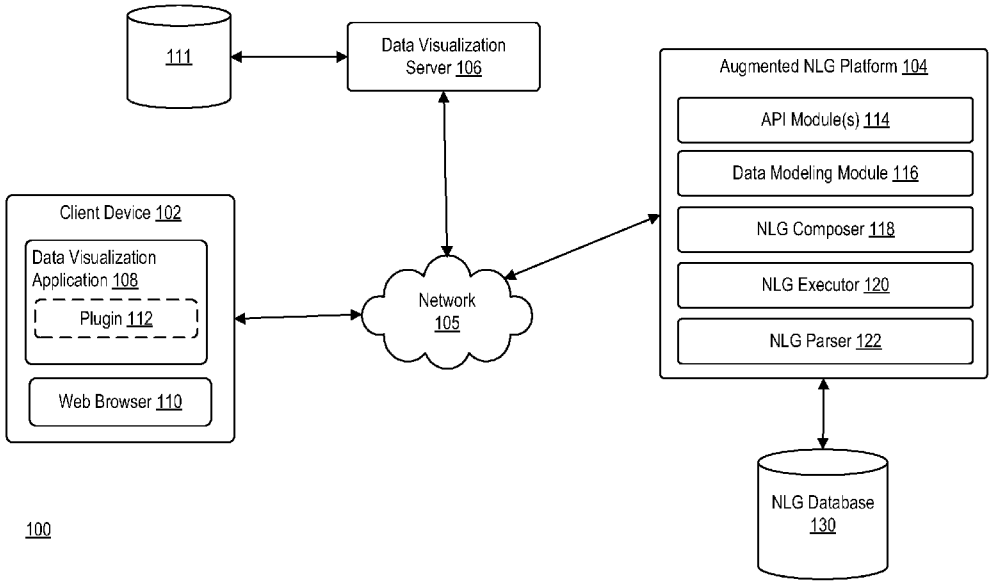


FIG. 1

(57) Abstract: A computing system receives a request to generate one or more narrative frameworks for a worksheet. The worksheet is hosted by a third party server. The computing system interfaces with the third party server to retrieve data associated with the worksheet. The computing system infers attributes of the worksheet from the retrieved data. The attributes correspond to a type of data represented in the worksheet. The computing system generates a narrative template for the one or more narrative frameworks. The computing system generates one or more data variables for the one or more narrative frameworks based on the generated narrative template. The computing system constructs the one or more narrative frameworks by selectively retrieving data from the worksheet for each data variable of the one or more data variables in the narrative framework.

WO 2021/067447 A1

[Continued on next page]

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))*

Augmented Natural Language Generation Platform

Cross-Reference to Related Applications

[0001] This application claims priority to U.S. Application Serial Number 62/908,235, filed September 30, 2019, which is hereby incorporated by reference in its entirety.

Field of the Disclosure

[0002] The present disclosure generally relates to an augmented natural language generation platform and a method of operating the same.

Background

[0003] The unprecedented rise of big data has revolutionized the intelligence industry, with trillions of gigabytes of data currently existing. In response to this influx of data, analysts and data scientists have created highly complex and congested dashboards to help analyze and interpret the massive amounts of data they typically have on hand. As a result, even the most experienced data scientists may struggle to adequately communicate the insights gained from the plurality of charts and graphs that were generated.

Summary

[0004] In some embodiments, a method is disclosed herein. A computing system receives a request to generate one or more narrative frameworks for a worksheet. The worksheet is hosted by a third party server. The computing system interfaces with the third party server to retrieve data associated with the worksheet. The computing system infers attributes of the worksheet from the retrieved data. The attributes correspond to a type of data represented in the worksheet. The computing system generates a narrative template for the one or more narrative frameworks. The computing system generates one or more data variables for the one or more narrative frameworks based on the generated narrative template. The computing system constructs the one or more narrative frameworks by selectively retrieving data from the worksheet for each data variable of the one or more data variables in the narrative framework.

[0005] In some embodiments, a non-transitory computer readable medium is disclosed herein. The non-transitory computer readable medium includes instructions which, when executed by a computing system, cause the computing system to perform operations. The operations include receiving, by a computing system, a request to generate one or more narrative frameworks for a worksheet. The worksheet is hosted by a third party server. The operations further include interfacing, by the computing system, with the third party server to retrieve data

associated with the worksheet. The operations further include inferring, by the computing system, attributes of the worksheet from the retrieved data. The attributes correspond to a type of data represented in the worksheet. The operations further include generating, by the computing system, a narrative template for the one or more narrative frameworks. The operations further include generating, by the computing system, one or more data variables for the one or more narrative frameworks based on the generated narrative template. The operations further include constructing, by the computing system, the one or more narrative frameworks by selectively retrieving data from the worksheet for each data variable of the one or more data variables in the narrative framework.

[0006] In some embodiments, a system is disclosed herein. The system includes a processor and a memory. The memory has programming instructions stored thereon, which, when executed by the processor, performs operations. The operations include receiving a request to generate one or more narrative frameworks for a worksheet. The worksheet is hosted by a third party server. The operations further include interfacing with the third party server to retrieve data associated with the worksheet. The operations further include inferring attributes of the worksheet from the retrieved data. The attributes correspond to a type of data represented in the worksheet. The operations further include generating a narrative template for the one or more narrative frameworks. The operations further include generating one or more data variables for the one or more narrative frameworks based on the generated narrative template. The operations further include constructing the one or more narrative frameworks by selectively retrieving data from the worksheet for each data variable of the one or more data variables in the narrative framework.

Brief Description of the Drawings

[0007] So that the manner in which the above recited features of the present disclosure can be understood in detail, a more particular description of the disclosure, briefly summarized above, may be had by reference to embodiments, some of which are illustrated in the appended drawings. It is to be noted, however, that the appended drawings illustrated only typical embodiments of this disclosure and are therefore not to be considered limiting of its scope, for the disclosure may admit to other equally effective embodiments.

[0008] Figure 1 is a block diagram illustrating a computing environment, according to example embodiments.

[0009] Figure 2A illustrates a graphical user interface corresponding to a step in the process flow of augmented natural language generation, according to example embodiments.

[0010] Figure 2B illustrates a graphical user interface corresponding to a step in the process flow of augmented natural language generation, according to example embodiments.

[0011] Figure 2C illustrates a graphical user interface corresponding to a step in the process flow of augmented natural language generation, according to example embodiments.

[0012] Figure 2D illustrates a graphical user interface corresponding to a step in the process flow of augmented natural language generation, according to example embodiments.

[0013] Figure 2E illustrates a graphical user interface corresponding to a step in the process flow of augmented natural language generation, according to example embodiments.

[0014] Figure 2F illustrates a graphical user interface corresponding to a step in the process flow of augmented natural language generation, according to example embodiments.

[0015] Figure 3 is a flow diagram illustrating a method of creating a report for a data sheet, according to example embodiments.

[0016] Figure 4A is a block diagram illustrating a computing device, according to example embodiments.

[0017] Figure 4B is a block diagram illustrating a computing device, according to example embodiments.

[0018] To facilitate understanding, identical reference numerals have been used, where possible, to designate identical elements that are common to the figures. It is contemplated that elements disclosed in one embodiment may be beneficially utilized on other embodiments without specific recitation.

Detailed Description

[0019] No matter how much data an organization can collect and process, at the end of the day, it is only as useful as the decisions that are gleaned from that data. Modern analytics dashboards and visualizations can help organize and arrange complex sets of data for interpretation. However, these dashboards are typically unable to generate actionable recommendations from the data. Extracting actual insights out of the various charts and graphs of a dashboard remains a very time extensive and subjective process.

[0020] One or more techniques described herein provides an augmented natural language generation engine for use with data visualization software. For example, one or more techniques described herein provide an augmented natural language generation engine that

works in conjunction with data visualization software to create highly targeted, dynamic, and compelling reports based on data managed by the data visualization software.

[0021] Figure 1 is a block diagram illustrating a computing environment 100, according to example embodiments. Computing environment 100 may include a client device 102, a augmented natural language generation platform 104, and an data visualization server 106 communicating via network 105.

[0022] Network 105 may be of any suitable type, including individual connections via the Internet, such as cellular or Wi-Fi networks. In some embodiments, network 105 may connect terminals, services, and mobile devices using direct connections, such as radio frequency identification (RFID), near-field communication (NFC), Bluetooth™, low-energy Bluetooth™ (BLE), Wi-Fi™, ZigBee™, ambient backscatter communication (ABC) protocols, USB, WAN, or LAN. Because the information transmitted may be personal or confidential, security concerns may dictate one or more of these types of connection be encrypted or otherwise secured. In some embodiments, however, the information being transmitted may be less personal, and therefore, the network connections may be selected for convenience over security.

[0023] Network 105 may include any type of computer networking arrangement used to exchange data or information. For example, network 105 may be the Internet, a private data network, virtual private network using a public network and/or other suitable connection(s) that enables components in computing environment 100 to send and receive information between the components of environment 100.

[0024] Client device 102 may be operated by a user. For example, client device 102 may be a mobile device, a tablet, a desktop computer, or any computing system having the capabilities described herein. Users may include, but are not limited to, individuals such as, for example, subscribers, clients, prospective clients, or customers of an entity associated with augmented natural language generation platform 104, such as individuals who have obtained, will obtain, or may obtain a product, service, or consultation from an entity associated with augmented natural language generation platform 104.

[0025] Client device 102 may include at least data visualization application 108. Data visualization application 108 may be representative of a stand-alone mobile or desktop data visualization software. Exemplary data visualization software may include, but is not limited to, Tableau, Sisense, Qlik, Plotly, Power BI, MicroStrategy, and the like. As illustrated, via data visualization application 108, client device 102 may access functionality of data

visualization server 106. For example, while interacting with data visualization application 108, a user may access various data sheets stored in database 111 by transmitting a request to data visualization server 106 via data visualization application 108.

[0026] Data visualization application 108 may include plugin 112. Plugin 112 may be associated with augmented natural language generation platform 104. For example, via plugin 112, client device 102 may access functionality of augmented natural language generation platform 104 directly within data visualization application 108. Client device 102 may communicate over network 105 to access functionality of augmented natural language generation platform 104. The content that is displayed to client device 102 may be transmitted from augmented natural language generation platform 104 to client device 102, and subsequently processed by data visualization application 108 for display through a graphical user interface (GUI) of client device 102.

[0027] In some embodiments, client device 102 may include web browser 110. Via web browser 110, client device 102 may access functionality of data visualization server 106 via one or more hypertext transfer protocol (HTTP) requests. In other words, in addition to accessing functionality of data visualization server 106 via a desktop or mobile application (e.g., data visualization application 108), a user may also access functionality of data visualization server 106 via a browser. In such embodiments, the browser-based data visualization application may include an integration with augmented natural language generation platform 104.

[0028] Augmented natural language generation platform 104 may include application programming interface (API) module 114, data modeling module 116, NLG composer 118, NLG executor 120, and NLG parser 122. Each or API module 114, data modeling module 116, NLG composer 118, NLG executor 120, and NLG parser 122 may be comprised of one or more software modules. The one or more software modules may be collections of code or instructions stored on a media (e.g., memory of augmented natural language generation platform 104) that represent a series of machine instructions (e.g., program code) that implement one or more algorithmic steps. Such machine instructions may be the actual computer code the processor of augmented natural language generation platform 104 interprets to implement the instructions or, alternatively, may be a higher level of coding of the instructions that is interpreted to obtain the actual computer code. The one or more software modules may also include one or more hardware components. One or more aspects of an

example algorithm may be performed by the hardware components (e.g., circuitry) itself, rather as a result of the instructions.

[0029] Augmented natural language generation platform 104 may be representative of a back-end complement to the front end integration of plugin 112 with data visualization application 108. Augmented natural language generation platform 104 may be configured to execute three functions: data modeling, etude processing, and etude execution. In operation, when a user interacts with plugin 112 via data visualization application 108, plugin 112 may prompt the user to select a target data set. The target data set may correspond to a data set associated with a user's account with the data visualization software. Upon receiving a selection of a target data set, plugin 112 may interface with augmented natural language generation platform 104. In some embodiments, plugin 112 may interface with augmented natural language generation platform 104 via one or more APIs managed by API module 114.

[0030] Upon receiving the target data set, data modeling module 116 may be configured to analyze the data set and extract information therefrom. In some embodiments, data modeling module 116 may be configured to infer data column types from the data set. In some embodiments, data modeling module 116 may be configured to infer detailed metadata for any columns and tables in the data set. This defines the type of data associated with the data set.

[0031] Generally, the data modeling performed by data modeling module 116 may be performed in two steps. During the first step, data modeling module 116 may attempt to cast a given string literal to an instance of an object with a certain type. If the cast is successful, the corresponding type may be returned. If however, the attempt fails, an error may be raised. In some embodiments, data modeling module 116 may loop over an ordered list of types until a successful cast is found. Data modeling module 116 may create a table schema based on the type inference. The table schema may be representative of a record of the identified types for each column.

[0032] The second step of the data modeling process may include three substeps: infer, aggregate, and discover. During the infer substep, data modeling module 116 may transform the rows of a table object into columns, with each column being assigned a subtype based on its type.

[0033] During the aggregate substep, data modeling module 116 may generate metrics based on the data. For example, data modeling module 116 may generate metrics, such, as but not limited to, count of null, sums based on unique records, totals, mean, median, standard deviation, and the like.

[0034] During the discover substep, data modeling module 116 may identify patterns among the table columns. To do so, data modeling module 116 may analyze the interactions between columns, instead of focusing on the individual columns themselves. In some embodiments, data modeling module 116 may identify hierarchies. Hierarchies may correspond to a series of one-to-many relationships among the columns of the table. These relationships may be represented by a directed graph, with keys parents and children. In some embodiments, this analysis may include columns with the string subtype. In some embodiments, data modeling module 116 may identify correlations. For example, data modeling module 116 may be configured to identify pairs of columns that are correlated with each other. In some embodiments, data modeling module 116 may be configured to identify pairs of columns that are perfectly correlated with each other. This analysis may be performed on all columns.

[0035] NLG composer 118 may be configured to analyze the inferred data generated by data modeling module 116 and may match the inferred data to an existing collection of motifs. A motif may refer to a blueprint for creating a template for the data. The blueprint may include the types of nodes that may be generated based on the data. In some embodiments, NLG composer 118 may match the inferred data to an existing set of motifs based on a specified mapping from motif fields to inputs to return a note, i.e., a collection of templates. In some embodiments, NLG composer 118 may recursively traverse through a compute graph to make an appropriate substitution in the merge fields. In some embodiments, such process can be specified by the fields themselves or overridden by various configurations. NLG executor 120 may be configured to analyze the mapped set of motifs and generate a project and template for the data. For example, NLG executor 120 may create various data variables and a template based on the mapped set of motifs.

[0036] NLG parser 122 may be configured to generate an executed report based on the data variables and the template generated by NLG executor 120. For example, based on the data variables and the template, NLG parser 122 may be configured to generate an actionable insight or note based for the data set. In some embodiments, NLG parser 122 may leverage a library to create a grammar and corresponding parser. In some embodiments, NLG parser 122 may parse the code into a structured abstract syntax tree (AST). As the AST is being built, NLG parser 122 may check the AST for static errors (e.g., invalid data types passed to a function/operator). In some embodiments, the AST may be evaluated to the result of the code.

[0037] Augmented natural language generation platform 104 may be configured to communicate with NLG database 130. NLG database 130 may be configured to store various motifs and templates for data processing.

[0038] Figure 2A illustrates an example view of a graphical user interface 202 corresponding to a step in the process flow of augmented natural language generation, according to example embodiments. Graphical user interface 202 (hereinafter "GUI 202") may correspond to a GUI presented to a user via client device 102. In some embodiments, GUI 202 may be a web page presented in web browser 110. In some embodiments, GUI 202 may be a graphical user interface generated by data visualization application 108. Generally, GUI 202 may correspond to an interface associated with plugin 112 functionality.

[0039] As illustrated, GUI 202 may correspond to an initial log-in page for plugin 112. In some embodiments, log-in page may allow a user of client device 102 to login to their account with augmented natural language generation platform 104. For example, via one or more fields 204, a user may enter their username (e.g., email) and password and be granted access to functionality of augmented natural language generation platform 104. In some embodiments, a user of client device 102 may not have an account with augmented natural language generation platform 104. In such embodiments, a user may interact with graphical element 206, which, when actuated, prompts the user to create an account with natural language generation platform.

[0040] Figure 2B illustrates an example view of a graphical user interface 212 corresponding to a step in the process flow of augmented natural language generation, according to example embodiments. Graphical user interface 212 (hereinafter "GUI 212") may correspond to a GUI presented to a user via client device 102. In some embodiments, GUI 202 may be a web page presented in web browser 110. In some embodiments, GUI 202 may be a graphical user interface generated by data visualization application 108. Generally, GUI 202 may correspond to an interface associated with plugin 112 functionality.

[0041] As illustrated, GUI 212 may be presented to a user following authentication with augmented natural language generation platform 104. Via GUI 212, a user may be presented with possible worksheets for augmented language generation. For example, following authentication with augmented natural language generation platform 104, plugin 112 may access data associated with the user and managed by data visualization server 106. In some embodiments, plugin 112 may access user data by leveraging one or more APIs of data visualization server 106. For example, via one or more APIs of data visualization server 106,

plugin 112 may be granted to access to a user's data stored in database 111. In some embodiments, the data stored in database 111 may correspond to one or more worksheets.

[0042] As shown, GUI 212 may include section 214. Section 214 may prompt the user to select a worksheet among a plurality of worksheets for natural language generation. In some embodiments, each worksheet in section 214 may include a graphical element 218 corresponding therefore. Via interaction with graphical element 218, a user may be navigated to the worksheet for review. In some embodiments, following interaction with graphical element 218, the user may be presented with an overlay window that includes a preview of the underlying worksheet.

[0043] Following selection of a worksheet for natural language generation, the user may interact with graphical element 216 to advance in the workflow.

[0044] Figure 2C illustrates an example view of a graphical user interface 222 corresponding to a step in the process flow of augmented natural language generation, according to example embodiments. Graphical user interface 222 (hereinafter "GUI 222") may correspond to a GUI presented to a user via client device 102. In some embodiments, GUI 222 may be a web page presented in web browser 1120. In some embodiments, GUI 222 may be a graphical user interface generated by data visualization application 108. GUI 222 may correspond to an interface associated with plugin 112 functionality.

[0045] As illustrated, GUI 222 may provide the user with possible notes to be added to a user's dashboard. GUI 222 may include one or more graphical elements 224-230. Graphical element 224 may correspond to a search bar. Via graphical element 224, a user may search for possible notes based on the data provided in the worksheet. For example, in operation, following selection of a worksheet via GUI 222, augmented natural language generation platform 104 may receive access to the worksheet via one or more APIs. Once augmented natural language generation platform 104 has access to the worksheet, data modeling module 116 may analyze the worksheet and extract information for further processing. For example, data modeling module 116 may parse the worksheet and infer data column types from the worksheet. In some embodiments, data modeling module 116 may further infer detailed metadata for any columns and tables in the data set. Data modeling module 116 may provide this information to NLG composer 118

[0046] NLG composer 118 may analyze the inferred data and may match the inferred data to an existing collection of motifs to create a template for the data. NLG executor 120 may receive the matching information from NLG composer 118 and generate a note and template for the

data. For example, NLG executor 120 may create various data variables and a template based on the mapped set of motifs. NLG executor 120 may provide the data variables and template to NLG parser for compilation.

[0047] NLG parser 122 may generate one or more notes 226, 228, and 230 based on the data variables and the template generated by NLG executor 120. In some embodiments, based on the data variables and the template, NLG parser 122 may generate an actionable insight or note based on data in the worksheet. For example, NLG parser 122 may selectively pull data from the worksheet based on which data corresponds to the data variables defined by NLG executor 120.

[0048] As shown, NLG parser 122 may generate a plurality of initial options to present to a user. Note 226 may correspond to a first option – "Average Largest Top 3." Based on the selected worksheet, NLG parser 122 may generate a note that reads: "Consumer is the Segment with the highest average SUM(Profit) at 1327.76, following by Corporate at 1022.76 and Home Office at 613.95." Note 228 may correspond to a second option – "Average Largest." Based on the selected worksheet, NLG parser 122 may generate a note that reads: "Consumer is the Segment with the largest SUM(Profit) at 1327.76." Note 230 may correspond to a third option – "Max Largest Last Month." Based on the selected worksheet, NLG parser 122 may generate a note that reads "The Segment with the largest total SUM(Profit) last month was Consumer with 4595.00." Each note may include an actionable element corresponding thereto. For example, as shown, actionable element 232 may correspond to note 226; actionable element 234 may correspond to note 228; and actionable element 236 may correspond to note 230. Via interaction with a respective actionable element 232-236, a user may select a respective note for display.

[0049] Figure 2D illustrates an example view of a graphical user interface 242 corresponding to a step in the process flow of augmented natural language generation, according to example embodiments. Graphical user interface 242 (hereinafter "GUI 242") may correspond to a GUI presented to a user via client device 102. In some embodiments, GUI 242 may be a web page presented in web browser 1120. In some embodiments, GUI 242 may be a graphical user interface generated by data visualization application 108. GUI 242 may correspond to an interface associated with plugin 112 functionality.

[0050] As illustrated, GUI 242 may allow a user to edit the generated note. GUI 242 may include actionable item 244. Actionable item 244 may correspond to a narrative framework of the selected note from GUI 232. For example, the representation of selected note may be

composed of one or more phrase variations 246. Each phrase variation may correspond to an item of data from the selected worksheet. Each phrase variation 246 may be actionable. For example, upon interacting with a phrase variation 246, a user may be presented with an overlay or pop-up window 245. Overlay or pop-up window 245 may include one or more format options for the selected phrase variation 246. For example, as shown, the user may interact with "613" to vary the way the data in phrase variation 246 is selected. In some embodiments, the user may add a dollar symbol in front of the number, define the number of decimal places, indicate the type of symbol used as a decimal separator, indicate if trailing zeros should be stripped, indicate if the value should be multiplied by 100, and the like. Once the note is tailored to the user's liking, a user may interact with actionable element 250 to accept the changes.

[0051] Figure 2E illustrates an example view of a graphical user interface 252 corresponding to a step in the process flow of augmented natural language generation, according to example embodiments. Graphical user interface 252 (hereinafter "GUI 252") may correspond to a GUI presented to a user via client device 102. In some embodiments, GUI 252 may be a web page presented in web browser 110. In some embodiments, GUI 252 may be a graphical user interface generated by data visualization application 108. GUI 252 may correspond to an interface associated with plugin 112 functionality.

[0052] As illustrated, GUI 252 may include graphical element 254. Graphical element 254 may include a representation of the approved note based on the narrative framework approved in GUI 242. GUI 252 may further include actionable element 256 and actionable element 258. Actionable element 256 may prompt the user to configurate the integration. Actionable element 258 may prompt the user to build another note. For example, following interaction with actionable element 258, a user may return to GUI 222 for selecting another note.

[0053] Figure 2F illustrates an example view of a graphical user interface 262 corresponding to a step in the process flow of augmented natural language generation, according to example embodiments. Graphical user interface 262 (hereinafter "GUI 262") may correspond to a GUI presented to a user via client device 102. In some embodiments, GUI 262 may be a web page presented in web browser 110. In some embodiments, GUI 262 may be a graphical user interface generated by data visualization application 108. GUI 262 may correspond to an interface associated with plugin 112 functionality.

[0054] As illustrated, GUI 262 may include graphical element 264 and graphical element 266. Graphical element 264 may allow a user to name their integration. For example, the user can

name the integration so that the user can refer back to the integration should the user wish to edit or modify the narrative framework. Graphical element 266 may allow a user to add additional worksheets to the integration. For example, as shown, graphical element 266 may include one or more prompts seeking user input to add additional worksheets. To complete the integration, a user may interact with graphical element 268.

[0055] Figure 3 is a flow diagram illustrating a method 300 of generating a narrative framework, according to example embodiments. Method 300 may begin at step 302.

[0056] At step 302, augmented natural language generation platform 104 may receive a request to generate one or more narrative frameworks for a worksheet. In some embodiments, augmented natural language generation platform 104 may receive a request from a plugin 112 executing with data visualization application 108 on client device 102.

[0057] At step 304, augmented natural language generation platform 104 may retrieve data corresponding to the worksheet from data visualization server 106. For example, augmented natural language generation platform 104 may perform an API call to retrieve the identified worksheet.

[0058] At step 306, augmented natural language generation platform 104 may identify one or more attributes of the worksheet. In some embodiments, data modeling module 116 may analyze the worksheet and extract information for further processing. For example, data modeling module 116 may parse the worksheet and infer data column types from the worksheet. In some embodiments, data modeling module 116 may further infer detailed metadata for any columns and tables in the data set. Data modeling module 116 may provide this information to NLG composer 118.

[0059] At step 308, augmented natural language generation platform 104 may create a narrative template for the worksheet, based on the attributes of the worksheet. In some embodiments, NLG composer 118 may analyze the inferred data. Using the inferred data, NLG composer may match the inferred data to an existing collection of motifs to create a template for the data. NLG composer 118 may provide this information to NLG executor 120.

[0060] At step 310, augmented natural language generation platform 104 may generate one or more data variables for the note. For example, NLG executor 120 may receive the matching information from NLG composer 118 and generate one or more data variables and template for the data. For example, NLG executor 120 may create various data variables and a template based on the mapped set of motifs. NLG executor 120 may provide the data variables and template to NLG parser for compilation.

[0061] At step 312, augmented natural language generation platform 104 may generate a finished note based on the one or more data variables and the template. For example, NLG parser 122 may generate narrative framework based on the data variables and the template generated by NLG executor 120. In some embodiments, based on the data variables and the template, NLG parser 122 may generate a narrative framework for the note based on data in the worksheet. For example, NLG parser 122 may selectively pull data from the worksheet based on which data corresponds to the data variables defined by NLG executor 120.

[0062] At step 314, augmented natural language generation platform 104 may associate the narrative framework with the worksheet. For example, if data in the selected worksheet changes over time, NLG executor 120 may be configured to update the values of the data variables based on the change. In some embodiments, when the notes are generated, a corresponding integration and project may be created and stored in database 130. For example, when worksheet data changes, the new data may be passed through the corresponding integration to update data variables in the notes.

[0063] Figure 4A illustrates a system bus computing system architecture 400, according to example embodiments. System 400 may be representative of at least a portion of augmented natural language generation platform 104. One or more components of system 400 may be in electrical communication with each other using a bus 405. System 400 may include a processing unit (CPU or processor) 410 and a system bus 405 that couples various system components including the system memory 415, such as read only memory (ROM) 420 and random access memory (RAM) 425, to processor 410. System 400 may include a cache of high-speed memory connected directly with, in close proximity to, or integrated as part of processor 410. System 400 may copy data from memory 415 and/or storage device 430 to cache 412 for quick access by processor 410. In this way, cache 412 may provide a performance boost that avoids processor 410 delays while waiting for data. These and other modules may control or be configured to control processor 410 to perform various actions. Other system memory 415 may be available for use as well. Memory 415 may include multiple different types of memory with different performance characteristics. Processor 410 may include any general-purpose processor and a hardware module or software module, such as service 1 432, service 2 434, and service 3 436 stored in storage device 430, configured to control processor 410 as well as a special-purpose processor where software instructions are incorporated into the actual processor design. Processor 410 may essentially be a completely

self-contained computing system, containing multiple cores or processors, a bus, memory controller, cache, etc. A multi-core processor may be symmetric or asymmetric.

[0064] To enable user interaction with the computing device 400, an input device 445 may represent any number of input mechanisms, such as a microphone for speech, a touch-sensitive screen for gesture or graphical input, keyboard, mouse, motion input, speech and so forth. An output device 435 may also be one or more of a number of output mechanisms known to those of skill in the art. In some instances, multimodal systems may enable a user to provide multiple types of input to communicate with computing device 400. Communications interface 440 may generally govern and manage the user input and system output. There is no restriction on operating on any particular hardware arrangement and therefore the basic features here may easily be substituted for improved hardware or firmware arrangements as they are developed.

[0065] Storage device 430 may be a non-volatile memory and may be a hard disk or other types of computer readable media which may store data that are accessible by a computer, such as magnetic cassettes, flash memory cards, solid state memory devices, digital versatile disks, cartridges, random access memories (RAMs) 425, read only memory (ROM) 420, and hybrids thereof.

[0066] Storage device 430 may include services 432, 434, and 436 for controlling the processor 410. Other hardware or software modules are contemplated. Storage device 430 may be connected to system bus 405. In one aspect, a hardware module that performs a particular function may include the software component stored in a computer-readable medium in connection with the necessary hardware components, such as processor 410, bus 405, display 435, and so forth, to carry out the function.

[0067] Figure 4B illustrates a computer system 450 having a chipset architecture that may represent at least a portion of augmented natural language generation platform 104. Computer system 450 may be an example of computer hardware, software, and firmware that may be used to implement the disclosed technology. System 450 may include a processor 455, representative of any number of physically and/or logically distinct resources capable of executing software, firmware, and hardware configured to perform identified computations. Processor 455 may communicate with a chipset 460 that may control input to and output from processor 455. In this example, chipset 460 outputs information to output 465, such as a display, and may read and write information to storage device 470, which may include magnetic media, and solid state media, for example. Chipset 460 may also read data from and write data to RAM 475. A bridge 480 for interfacing with a variety of user interface

components 485 may be provided for interfacing with chipset 460. Such user interface components 485 may include a keyboard, a microphone, touch detection and processing circuitry, a pointing device, such as a mouse, and so on. In general, inputs to system 450 may come from any of a variety of sources, machine generated and/or human generated.

[0068] Chipset 460 may also interface with one or more communication interfaces 490 that may have different physical interfaces. Such communication interfaces may include interfaces for wired and wireless local area networks, for broadband wireless networks, as well as personal area networks. Some applications of the methods for generating, displaying, and using the GUI disclosed herein may include receiving ordered datasets over the physical interface or be generated by the machine itself by processor 455 analyzing data stored in storage 470 or 475. Further, the machine may receive inputs from a user through user interface components 485 and execute appropriate functions, such as browsing functions by interpreting these inputs using processor 455.

[0069] It may be appreciated that example systems 400 and 450 may have more than one processor 410 or be part of a group or cluster of computing devices networked together to provide greater processing capability.

[0070] While the foregoing is directed to embodiments described herein, other and further embodiments may be devised without departing from the basic scope thereof. For example, aspects of the present disclosure may be implemented in hardware or software or a combination of hardware and software. One embodiment described herein may be implemented as a program product for use with a computer system. The program(s) of the program product define functions of the embodiments (including the methods described herein) and can be contained on a variety of computer-readable storage media. Illustrative computer-readable storage media include, but are not limited to: (i) non-writable storage media (e.g., read-only memory (ROM) devices within a computer, such as CD-ROM disks readably by a CD-ROM drive, flash memory, ROM chips, or any type of solid-state non-volatile memory) on which information is permanently stored; and (ii) writable storage media (e.g., floppy disks within a diskette drive or hard-disk drive or any type of solid state random-access memory) on which alterable information is stored. Such computer-readable storage media, when carrying computer-readable instructions that direct the functions of the disclosed embodiments, are embodiments of the present disclosure.

[0071] It will be appreciated to those skilled in the art that the preceding examples are exemplary and not limiting. It is intended that all permutations, enhancements, equivalents,

and improvements thereto are apparent to those skilled in the art upon a reading of the specification and a study of the drawings are included within the true spirit and scope of the present disclosure. It is therefore intended that the following appended claims include all such modifications, permutations, and equivalents as fall within the true spirit and scope of these teachings.

Claims:

1. A method, comprising:

receiving, by a computing system, a request to generate one or more narrative frameworks for a worksheet, the worksheet hosted by a third party server;

interfacing, by the computing system, with the third party server to retrieve data associated with the worksheet;

inferring, by the computing system, attributes of the worksheet from the retrieved data, the attributes corresponding to a type of data represented in the worksheet;

generating, by the computing system, a narrative template for the one or more narrative frameworks;

generating, by the computing system, one or more data variables for the one or more narrative frameworks based on the generated narrative template; and

constructing, by the computing system, the one or more narrative frameworks by selectively retrieving data from the worksheet for each data variable of the one or more data variables in the narrative framework.

2. The method of claim 1, wherein the request to generate the one or more narrative frameworks is received from a plugin executing with a data visualization application on a client device.

3. The method of claim 1, wherein interfacing, by the computing system, with the third party server to retrieve data associated with the worksheet comprises interfacing with the third party server via one or more application programming interfaces (APIs) linking the computing system to the third party server.

4. The method of claim 1, further comprising:

receiving, by the computing system, a request to modify a value of a data variable of the one or more data variables in the narrative framework; and

updating, by the computing system, the data variable of the one or more data variables based on the request.

5. The method of claim 1, further comprising:

receiving, by the computing system from the third party server, an update to the worksheet;

identifying, by the computing system, that the update to the worksheet changes a value of a data variable of the one or more data variables in the narrative framework; and

updating, by the computing system, the narrative framework based on the update to the worksheet.

6. The method of claim 1, wherein inferring, by the computing system, the attributes of the worksheet from the retrieved data comprises:

inferring a data type associated with each column of data in the worksheet.

7. The method of claim 1, wherein generating, by the computing system, the narrative template for the one or more narrative frameworks comprises:

matching the inferred attributes to a collection of motifs.

8. A non-transitory computer readable medium comprising instructions which, when executed by a computing system, cause the computing system to perform operations comprising:

receiving, by a computing system, a request to generate one or more narrative frameworks for a worksheet, the worksheet hosted by a third party server;

interfacing, by the computing system, with the third party server to retrieve data associated with the worksheet;

inferring, by the computing system, attributes of the worksheet from the retrieved data, the attributes corresponding to a type of data represented in the worksheet;

generating, by the computing system, a narrative template for the one or more narrative frameworks;

generating, by the computing system, one or more data variables for the one or more narrative frameworks based on the generated narrative template; and

constructing, by the computing system, the one or more narrative frameworks by selectively retrieving data from the worksheet for each data variable of the one or more data variables in the narrative framework.

9. The non-transitory computer readable medium of claim 8, wherein the request to generate the one or more narrative frameworks is received from a plugin executing with a data visualization application on a client device.

10. The non-transitory computer readable medium of claim 8, wherein interfacing, by the computing system, with the third party server to retrieve data associated with the worksheet comprises interfacing with the third party server via one or more application programming interfaces (APIs) linking the computing system to the third party server.

11. The non-transitory computer readable medium of claim 8, further comprising:

receiving, by the computing system, a request to modify a value of a data variable of the one or more data variables in the narrative framework; and

updating, by the computing system, the data variable of the one or more data variables based on the request.

12. The non-transitory computer readable medium of claim 8, further comprising:

receiving, by the computing system from the third party server, an update to the worksheet;

identifying, by the computing system, that the update to the worksheet changes a value of a data variable of the one or more data variables in the narrative framework; and

updating, by the computing system, the narrative framework based on the update to the worksheet.

13. The non-transitory computer readable medium of claim 8, wherein inferring, by the computing system, the attributes of the worksheet from the retrieved data comprises:

inferring a data type associated with each column of data in the worksheet.

14. The non-transitory computer readable medium of claim 8, wherein generating, by the computing system, the narrative template for the one or more narrative frameworks comprises:

matching the inferred attributes to a collection of motifs.

15. A system comprising:

a processor; and

a memory having programming instructions stored thereon, which, when executed by the processor, performs operations comprising:

receiving a request to generate one or more narrative frameworks for a worksheet, the worksheet hosted by a third party server;

interfacing with the third party server to retrieve data associated with the worksheet;

inferring attributes of the worksheet from the retrieved data, the attributes corresponding to a type of data represented in the worksheet;

generating a narrative template for the one or more narrative frameworks;

generating one or more data variables for the one or more narrative frameworks based on the generated narrative template; and

constructing the one or more narrative frameworks by selectively retrieving data from the worksheet for each data variable of the one or more data variables in the narrative framework.

16. The system of claim 15, wherein the request to generate the one or more narrative frameworks is received from a plugin executing with a data visualization application on a client device.

17. The system of claim 15, wherein interfacing with the third party server to retrieve data associated with the worksheet comprises interfacing with the third party server via one or more application programming interfaces (APIs) linking the system to the third party server.

18. The system of claim 15, wherein the operations further comprise:

receiving a request to modify a value of a data variable of the one or more data variables in the narrative framework; and

updating the data variable of the one or more data variables based on the request.

19. The system of claim 15, wherein the operations further comprise:

receiving, from the third party server, an update to the worksheet;

identifying that the update to the worksheet changes a value of a data variable of the one or more data variables in the narrative framework; and

updating the narrative framework based on the update to the worksheet.

20. The system of claims 15, wherein inferring the attributes of the worksheet from the retrieved data comprises:

inferring a data type associated with each column of data in the worksheet.

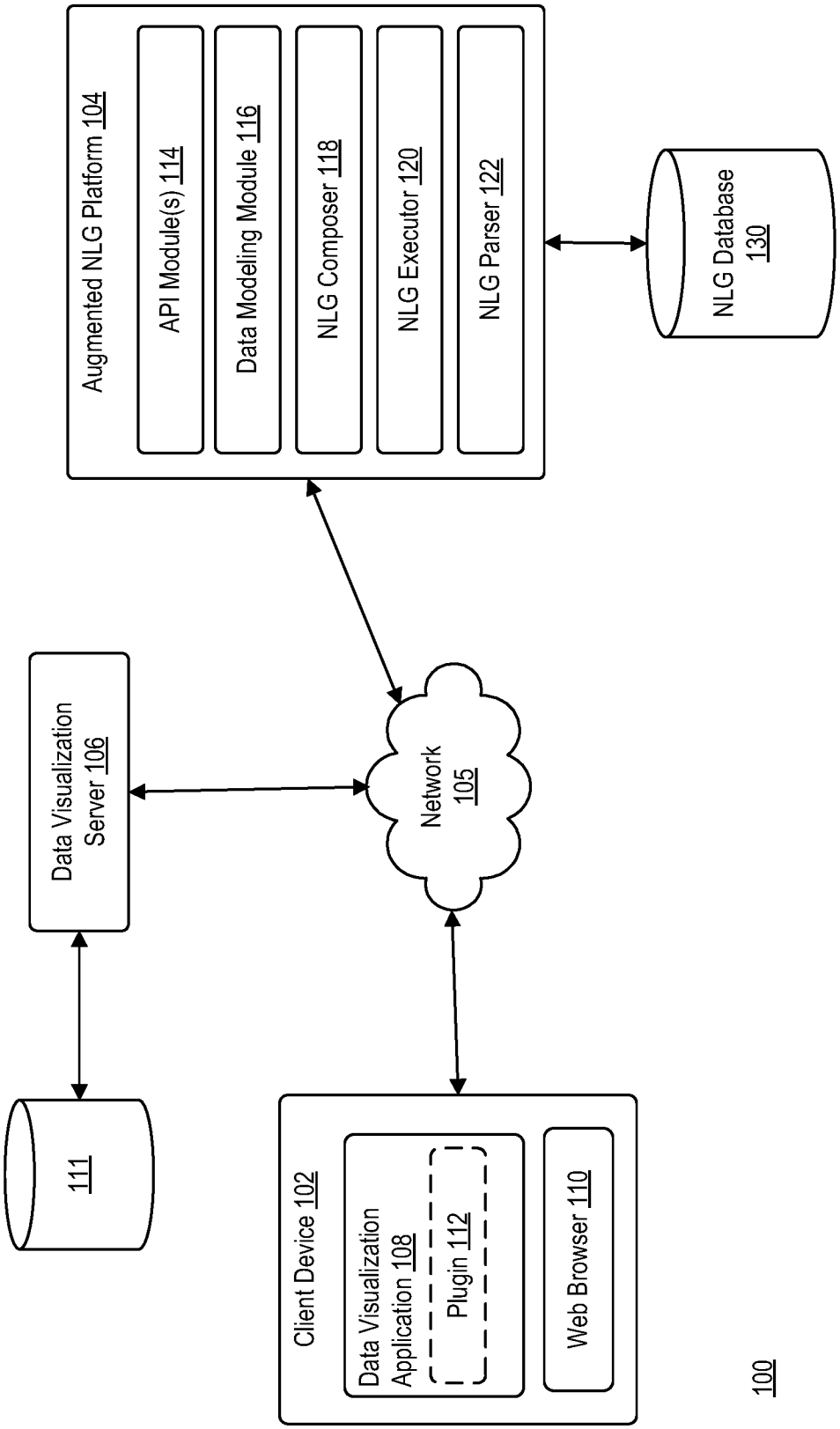

Augmented NLG Platform 104
API Module(s) 114
Data Modeling Module 116
NLG Composer 118
NLG Executor 120
NLG Parser 122
NLG Database 130
Data Visualization Server 106
Network 105
111
Client Device 102
Data Visualization Application 108
Plugin 112
Web Browser 110
100

FIG. 1

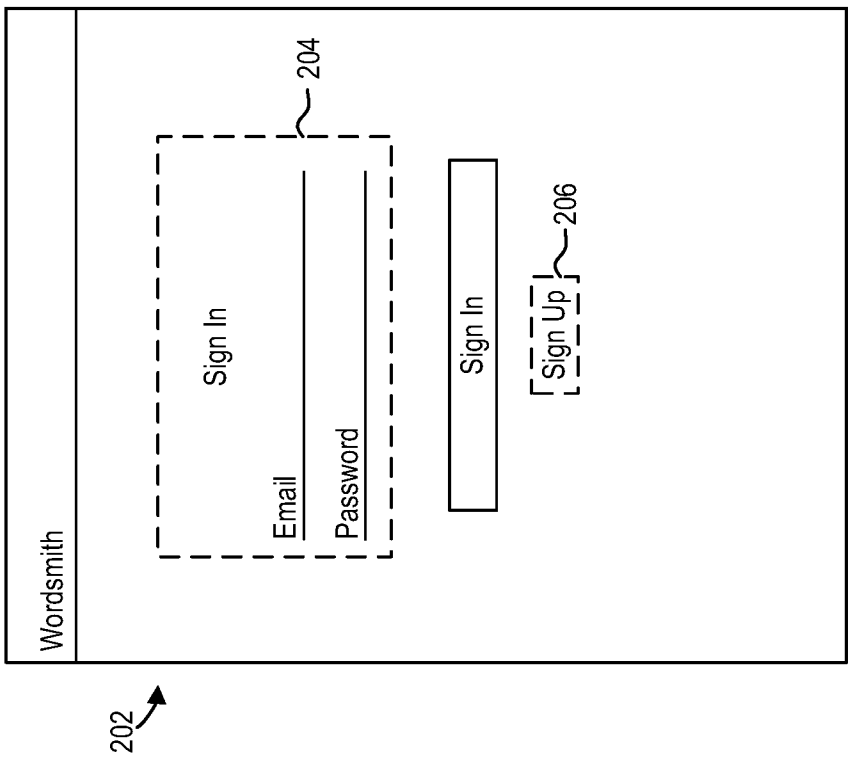

Wordsmith
Sign In
Email
Password
204
Sign In
Sign Up
206
202

FIG. 2A

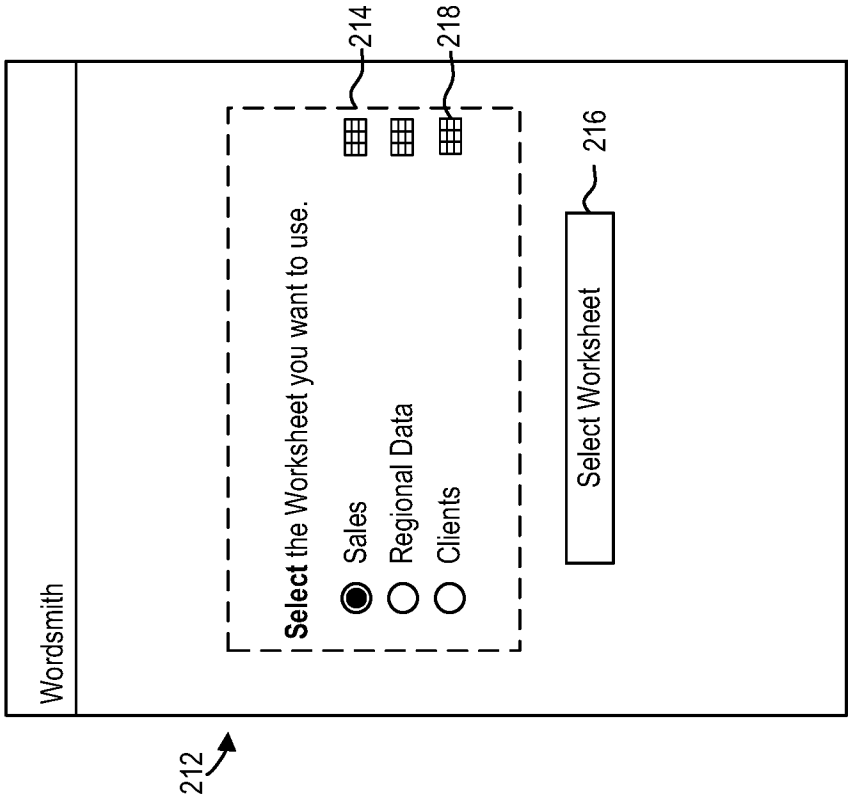

Wordsmith
Select the Worksheet you want to use.
Sales
Regional Data
Clients
214
218
Select Worksheet
216
212

FIG. 2B

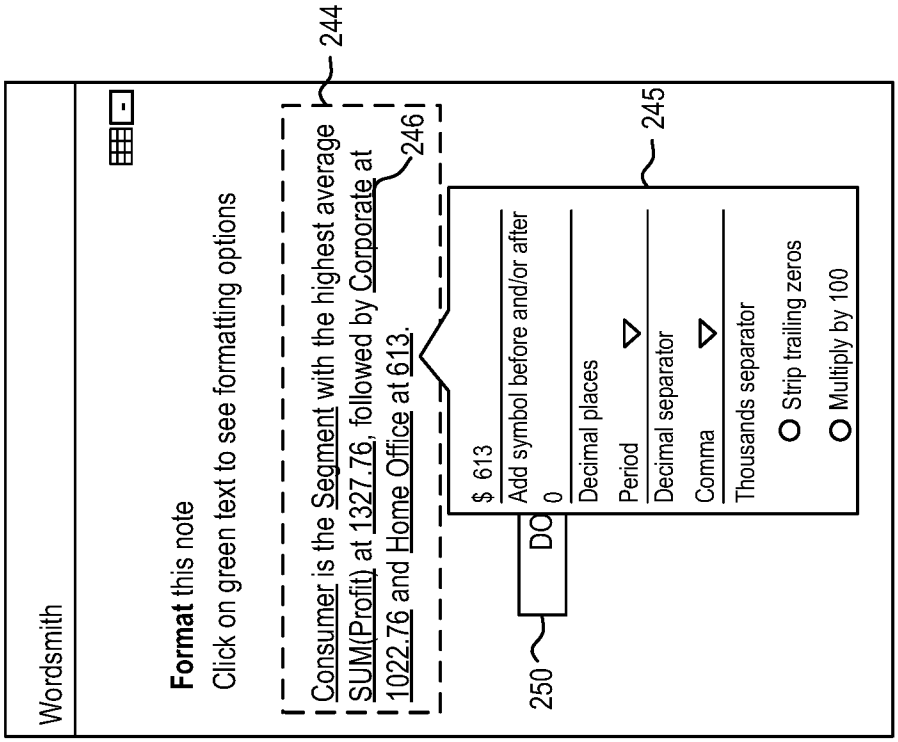

242
Wordsmith
Format this note
Click on green text to see formatting options
244
Consumer is the Segment with the highest average SUM(Profit) at 1327.76, followed by Corporate at 1022.76 and Home Office at 613.
246
$ 613
Add symbol before and/or after
0
Decimal places
Period
Decimal separator
Comma
Thousands separator
Strip trailing zeros
Multiply by 100
245
DO
250

FIG. 2D

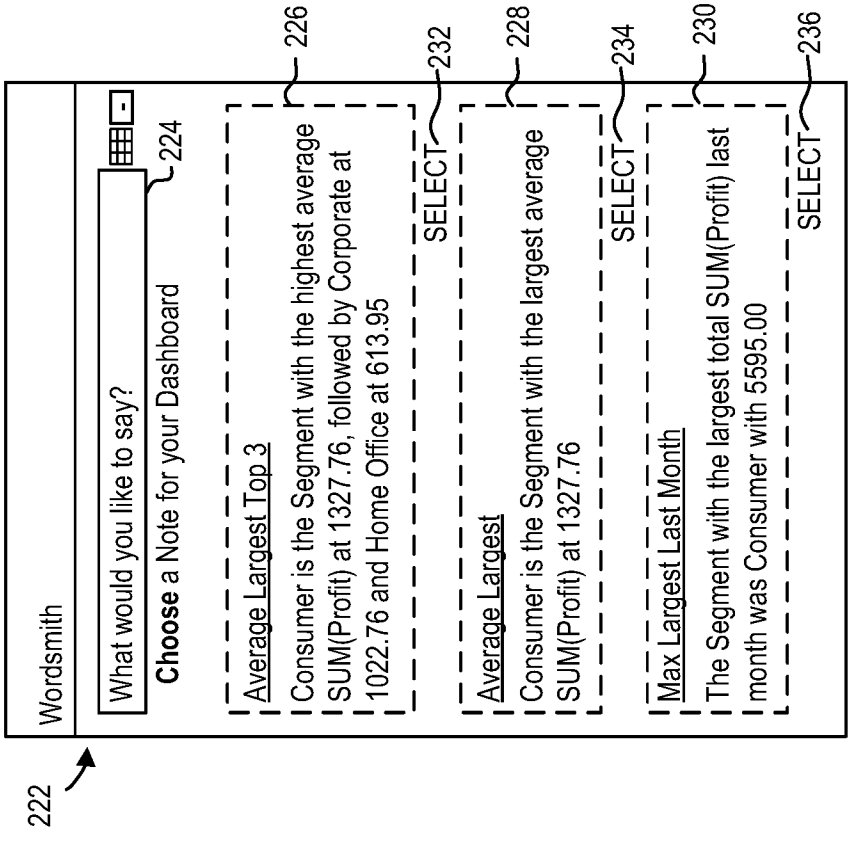

222
Wordsmith
What would you like to say?
Choose a Note for your Dashboard
224
Average Largest Top 3
Consumer is the Segment with the highest average SUM(Profit) at 1327.76, followed by Corporate at 1022.76 and Home Office at 613.95
226
SELECT
232
Average Largest
Consumer is the Segment with the largest average SUM(Profit) at 1327.76
228
SELECT
234
Max Largest Last Month
The Segment with the largest total SUM(Profit) last month was Consumer with 5595.00
230
SELECT
236

FIG. 2C

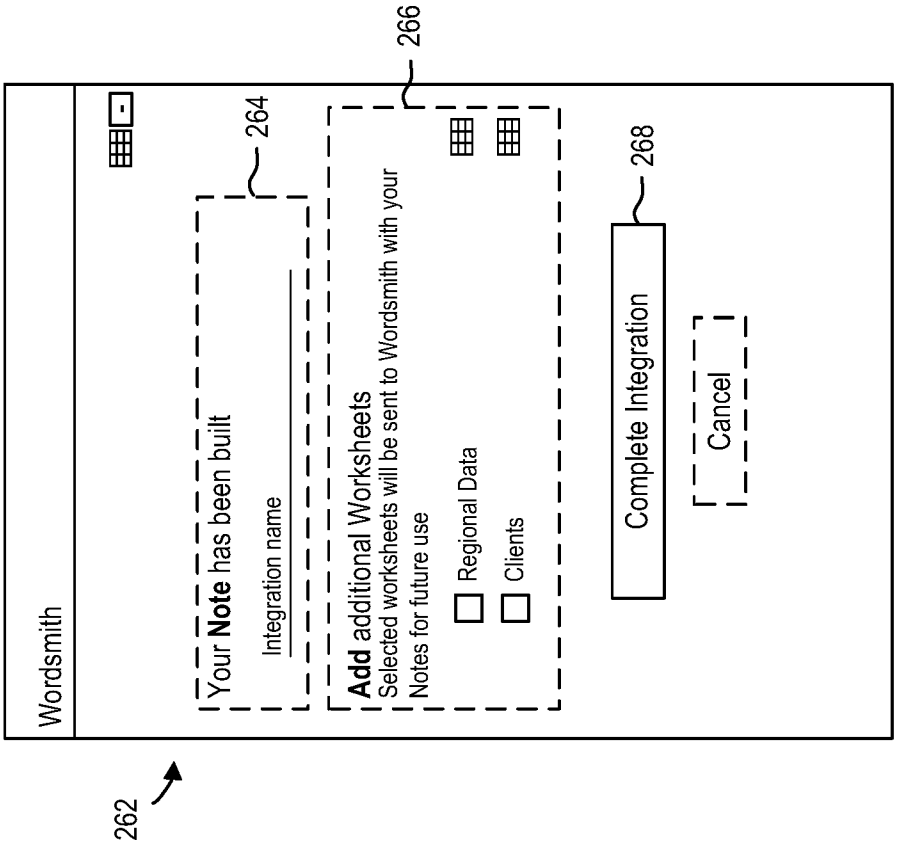

Wordsmith
Your Note has been built
Consumer is the Segment with the highest average SUM(Profit) at 1327.76, followed by Corporate at 1022.76 and Home Office at 613.
Configure Integration
Build another note
Cancel
252
254
256
258
258

FIG. 2E

Wordsmith
Your Note has been built
Integration name
Add additional Worksheets
Selected worksheets will be sent to Wordsmith with your Notes for future use
Regional Data
Clients
Complete Integration
Cancel
262
264
266
268

FIG. 2F

300

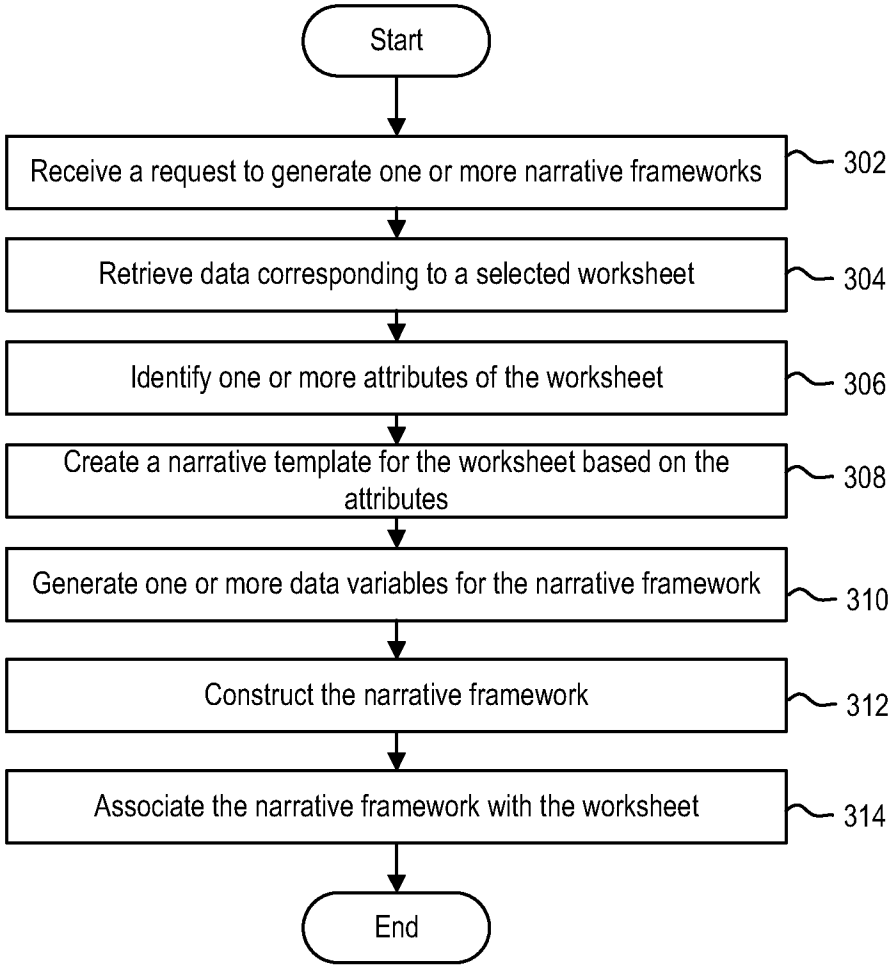

Start
Receive a request to generate one or more narrative frameworks
302
Retrieve data corresponding to a selected worksheet
304
Identify one or more attributes of the worksheet
306
Create a narrative template for the worksheet based on the attributes
308
Generate one or more data variables for the narrative framework
310
Construct the narrative framework
312
Associate the narrative framework with the worksheet
314
End

FIG. 3

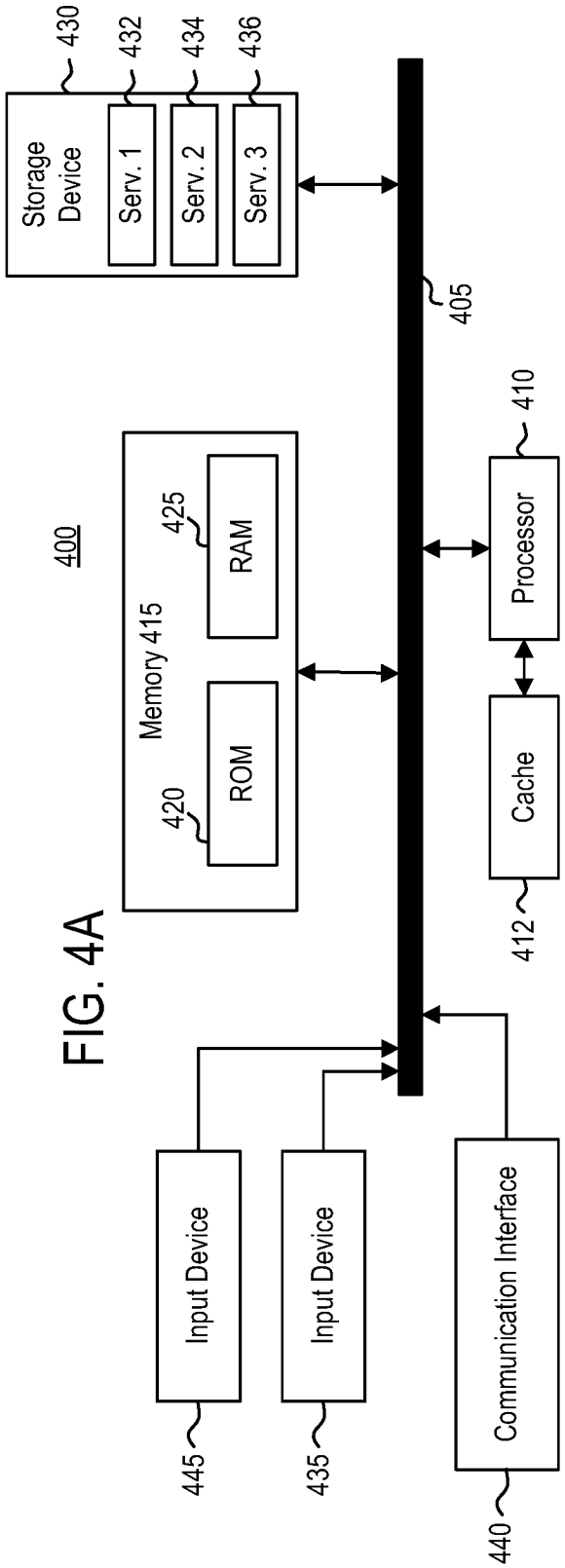

400
Storage Device
430
Serv. 1
432
Serv. 2
434
Serv. 3
436
405
Memory 415
425
RAM
420
ROM
Processor
410
Cache
412
Input Device
445
Input Device
435
Communication Interface
440

FIG. 4A

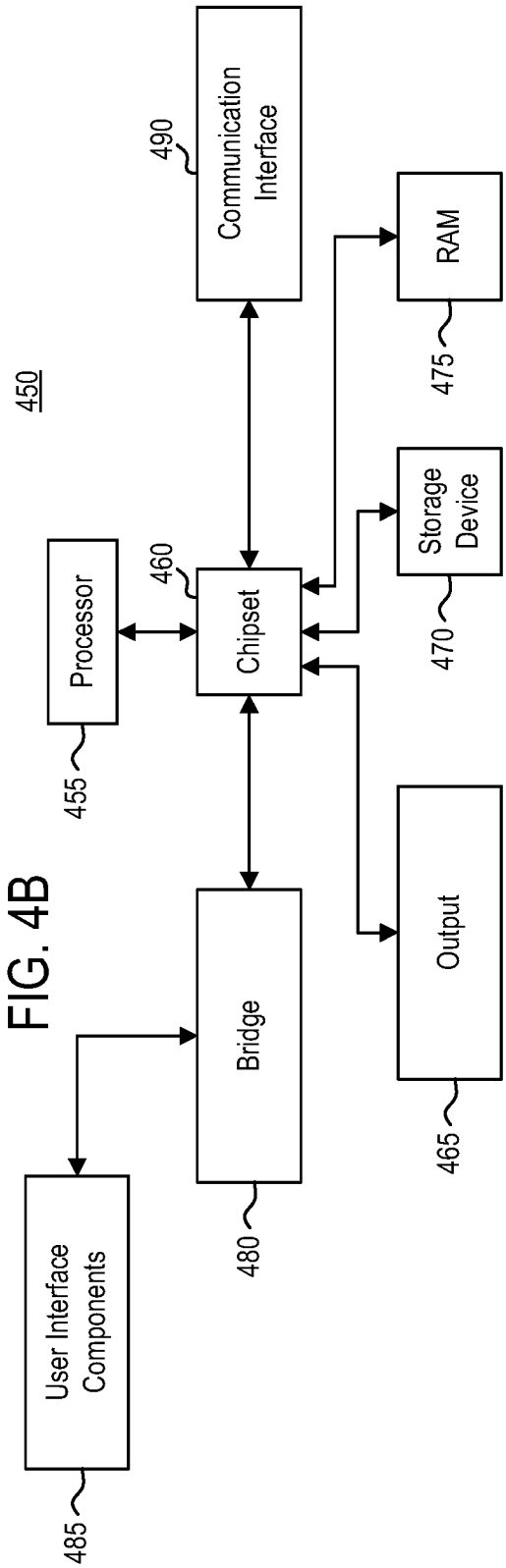

450
Communication Interface
490
RAM
475
Processor
455
460
Chipset
Storage Device
470
Bridge
480
Output
465
User Interface Components
485

FIG. 4B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US20/53560

A. CLASSIFICATION OF SUBJECT MATTER

IPC - G06F 40/289; G06F 40/35; G06F 40/56; G10L 13/04 (2020.01)

CPC - G06F 40/289; G06F 40/56; G06F 40/211; G06F 40/35; G06F 40/10; G06F 40/186; G10L 13/04; G06F 40/106; G06F 16/3344; G06F 16/338

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 --- Y	US 2018/0329987 A1 (ACCENTURE GLOBAL SOLUTIONS LIMITED) 15 November 2018; figures 9A, 12; paragraphs [0024], [0031], [0050], [0071], [0073]; claims 1, 18	1-4, 6, 8-11, 13, 15-18, 20 --- 5, 7, 12, 14, 19
Y	US 2016/0078017 A1 (AUTOMATED INSIGHTS, INC) 17 March 2016; abstract; figures 11; paragraphs [0030], [0033], [0047]	5, 12, 19
Y	WO 2017/003889 A1 (MICROSOFT TECHNOLOGY LICENSING, LLC) 05 January 2017; paragraphs [0012], [0019], [0065]	7, 14
A	US 2016/0086085 A1 (PRIMAL FUSION INC) 24 March 2016; 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
10 December 2020 (10.12.2020)	12 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 Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300