

(51) International Patent Classification:
G06F 17/00 (2006.01)(21) International Application Number:
PCT/US2014/036056(22) International Filing Date:
30 April 2014 (30.04.2014)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: HEWLETT-PACKARD DEVELOPMENT
COMPANY, L.P. [US/US]; 11445 Compaq Center Drive
W., Houston, Texas 77070 (US).(72) Inventors: HAO, Ming C.; 1501 Page Mill Road, Palo
Alto, California 94304-1100 (US). HUND, Michael; Uni-
versitätsstraße 10, 78457 Konstanz (DE). ROHRDANTZ,
Christian; Universitätsstraße 10, 48457 Konstanz (DE).
GHOSH, Riddhiman; 1501 Page Mill Road, Palo Alto,
California 94304-1100 (US). CHANG, Nelson L.; 1501
Page Mill Road, Palo Alto, California 94304-1100 (US).
HSU, Meichun; 1140 Enterprise Way, Palo Alto, Califor-
nia 94089 (US). KEIM, Daniel; Universitätsstraße 10,
78457 Konstanz (DE).(74) Agents: KIRCHEV, Ivan Tomov et al.; Hewlett-packard
Company, Intellectual Property Administration, Mail Stop
35 3404 E. Harmony Road, Fort Collins, Colorado 80528
(US).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM,
ZW.(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, 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 the identity of the inventor (Rule 4.17(i))
- 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))

(54) Title: VISUALIZING TOPICS WITH BUBBLES INCLUDING PIXELS

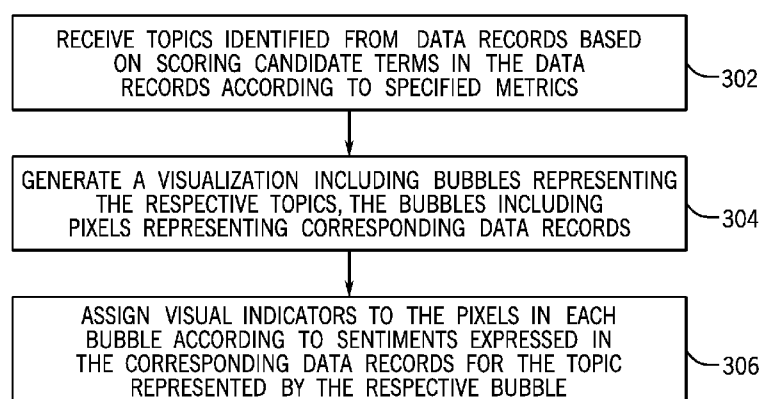


FIG. 3

(57) Abstract: Selected topics are identified from records based on scoring candidate terms in the records according to a user-specified metric and at least one further metric selected from among frequencies of occurrence of records pertaining to the respective candidate terms, and negativity of sentiment expressed with respect to the candidate terms in the records. A visualization is generated that includes bubbles representing the respective topics, the bubbles including pixels representing corresponding records, where a given one of the bubbles has a shape dependent upon a number of records represented by the given bubble and a time interval represented by the given bubble. Visual indicators are assigned to the pixels in the given bubble according to values of an attribute expressed in the corresponding records for the topic represented by the given bubble.

VISUALIZING TOPICS WITH BUBBLES INCLUDING PIXELS

Background

[0001] An enterprise can receive various types of user reviews that contain feedback about the enterprise. User reviews can be submitted online at third-party sites (*e.g.* web survey sites or social networking sites). Alternatively, reviews can be received directly by an enterprise. There can potentially be a relatively large number of data records containing user reviews, which can make meaningful analysis of such data records difficult and time-consuming.

Brief Description Of The Drawings

[0002] Some implementations are described with respect to the following figures.

[0003] Fig. 1 illustrates an example graphical visualization that includes bubbles representing respective topics, in accordance with some implementations.

[0004] Fig. 2 is an example schematic diagram of a bubble including pixels assigned to visual indicators corresponding to sentiments expressed in data records, in accordance with some implementations.

[0005] Fig. 3 is a flow diagram of an example visualization process at a client device, according to some implementations.

[0006] Fig. 4 is a flow diagram of example data record processing at a server device according to some implementations.

[0007] Fig. 5 is a flow diagram of an example process at a client device, according to some implementations.

[0008] Fig. 6 is a block diagram of an example arrangement including a client device and a server device, according to some implementations.

Detailed Description

[0009] An enterprise (*e.g.* a company, educational organization, government agency, individual, etc.) may collect feedback from users that can be used to better understand user sentiment regarding the enterprise, such as about an offering of the enterprise or another feature of the enterprise. An offering can include a product or a service provided by the enterprise. A “sentiment” refers to an attitude, opinion, or judgment of a human with respect to the offering. An opinion or other sentiment can be mapped to an attribute (such as an attribute associated with an offering) to indicate a degree of satisfaction or other sentiment with the attribute.

[0010] User feedback can be collected from various sources, such as online websites of the enterprise or third-party sites such as travel review websites, product review websites, social networking sites, web survey sites, customer support agents, and so forth. The user feedback can be received in data records. A “data record” can refer to a unit of data that contains collected information. For example, the data record can include a user review submitted by a particular user, in which the user may have expressed a sentiment with respect to an attribute. A data record can include one or multiple attributes, where an attribute can refer to an item (*e.g.* product offering, service offering, sport team, players of a sport team, a movie, a song, etc.) that may be the subject of review or feedback from users.

[0011] Visualizing relatively large volumes of data records containing user feedback for the purpose of sentiment analysis can be complex and time-consuming. In some examples, to discover issues that may be expressed by user feedback, an analysis of information contained in data records can be performed. For example, the analysis of the user feedback contained in the data records can attempt to identify time periods of increased negative sentiment, determine root causes of the negative sentiment, and to identify actions to take in response to the determined root causes.

[0012] Traditionally, the analysis of user feedback is performed manually, which can be time consuming and labor intensive. Also, manual analysis is prone to error, as an analyst may miss issues that may be expressed in data records. Also, manual analysis is usually performed in an offline manner based on a log of data records that was collected in the past. As a result, the identification of an issue may be delayed for a substantial amount of time.

[0013] In accordance with some implementations, automated visual analytics techniques or mechanisms are provided for identifying and visualizing issues in user feedback expressed in data records. In some implementations, the identification and visualization of issues in user feedback can be performed on a “real-time” basis, in which the identification and visualization can be performed as data records are being received. Although reference is made to real-time visual analytics in some implementations, it is noted that in alternative implementations, visual analytics can be performed in an offline manner on data records stored in a historical log or other historical data collection.

[0014] The identified issues are represented as topics, where a topic provides a description or indication of an issue expressed in user feedback. Once topics are identified, the topics are depicted in a visualization (in the form of a graphical representation) that includes bubbles containing pixels. Each topic is represented by a set of one or multiple bubbles in the visualization. A “bubble” can refer to a discrete region containing pixels in the visualization that is visibly distinct from another region containing pixels in the visualization. The pixels of a bubble represent respective data records (*e.g.* user reviews) within a respective time interval for a given topic. Multiple bubbles representing a given topic correspond to respective different time intervals.

[0015] A pixel can refer to a graphical element that can be displayed in a visualization, where the graphical element can be in the form of a dot, square, circle, or any other shape. Each pixel is assigned a visual indicator to indicate a sentiment expressed with respect to an attribute contained in a data record. Pixels are assigned different visual indicators to depict corresponding different sentiments expressed with respect to a respective topic. For example, the visual indicators can include different colors. A first color can be assigned to a pixel for a positive user sentiment expressed with respect to the corresponding attribute, while a second, different color can be assigned to represent a negative user sentiment expressed with respect to the attribute. In some examples, a positive sentiment can be indicated by a green color, while a negative sentiment can be indicated by a red color. Different shades of green can express different levels of positive user sentiment (*e.g.* darker green can express a more positive sentiment than a lighter green). Similarly, different shades of red can express different levels of negative sentiment (*e.g.* darker red can express a more negative sentiment

that lighter red). A neutral sentiment (a sentiment that is neither positive nor negative) can be expressed by a different color, such as a gray or white color. In other examples, other colors can be used to express different sentiments.

[0016] In some examples, the sentiment expressed in a particular data record for a given attribute can be based on scores assigned by a user (*e.g.*, a score between 1 and 5, where 5 is positive while 1 is negative and 3 is neutral, or a score from among -2, -1, 0, +1, +2, where the positive values reflect positive sentiment, negative values reflect negative sentiment, and 0 reflects a neutral sentiment). In other examples, the sentiment expressed in a particular data record for a given attribute can be based on an analysis of words in the particular data record. For example, the analysis can identify sentiment words associated with nouns (or compound nouns) that are targets of the sentiment words. The sentiment words can be processed to determine the corresponding sentiment (*e.g.* positive sentiment, negative sentiment, neutral sentiment).

[0017] Fig. 1 illustrates an example visualization 100 that has a first axis 102 and a second axis 104. Different points along the first axis 102 represents respective topics, while different points along the second axis 104 represents different points in time. The example topics depicted in the visualization 100 include “TEAM A,” “GAME,” “STREAK,” and so forth. Various bubbles are included in the visualization 100, with bubbles 106, 107, and 108 being referenced in Fig. 1. Each bubble includes an arrangement of pixels assigned respective different colors corresponding to sentiments expressed with respect to the respective topic. Different cross hatchings within each bubble represent respective different collections of pixels assigned corresponding different colors.

[0018] In the example of Fig. 1, the bubbles 106, 107, and 108 represent the “TEAM A” topic, and different colors assigned to the pixels in the bubbles 106, 107, and 108 are based on respective sentiments expressed with respect to the “TEAM A” topic for different respective time intervals. The topics 106, 107, and 108 are arranged across a row in the visualization 100 (each row in the visualization 100 is made up of a horizontal line or arrangement of one or multiple bubbles). Each different row in the visualization 100 includes one or multiple bubbles corresponding to a respective different topic. For the sake of clarity, some of the topics are not listed in the visualization of Fig. 1.

[0019] More generally, bubbles arranged along a given line (*e.g.* row) that is parallel to the time axis of the visualization 100 represent a respective topic.

[0020] In the example of Fig. 1, it is assumed that data records represented by pixels in the various bubbles of the visualization 100 include user feedback with respect to teams, players, and other attributes of a baseball world series. Fig. 1 also shows a group 110 of bubbles that represent various topics at the beginning of the world series, and a second group 112 of bubbles that represent topics at the end of the world series.

[0021] The visualization 100 can be an interactive visualization, such that a user can interact with elements in the visualization 100. For example, using a user input device (*e.g.* mouse device, touchscreen, keyboard, etc.), a user can move a cursor over a pixel included in a bubble in the visualization 100. Once the cursor is moved over the pixel, a dialog box can pop up, where the dialog box can include further details regarding the content of the data record (*e.g.* user review) represented by the pixel.

[0022] In addition, to zoom into a portion of the visualization, a user can select a bubble, such as by clicking on a button of a mouse device, tapping on the location of the bubble on a touchscreen, and so forth. In response to user selection of a bubble, another visualization screen can be displayed, such as visualization screen 114 shown in Fig. 1, which provides a zoomed view. In the example of Fig. 1, the bubble selected by the user can be bubble 107. In response to user selection of the bubble 107, the visualization screen 114 displays the selected bubble 107, along with other bubbles that are associated with the selected bubble 107. In the visualization screen 114, a bubble representing a “PLAYER A” topic is linked by a link 116 to the selected bubble 107. Also, the bubble representing the “PLAYER A” topic is linked by a link 116 to a bubble representing a “TEAM B” topic. Bubbles can be linked together if they share data records or information in data records, for example, or according to another criterion. Note that the bubbles in the visualization screen 114 are bubbles containing pixels representing data records that correspond to a specific time interval. In this example, bubbles from different time intervals are not linked together. The visualization screen 114 thus provides a zoomed view, and also links a selected bubble with other bubbles.

[0023] Fig. 2 is a schematic diagram of a bubble 200, according to some implementations. As shown in Fig. 2, pixels are depicted as little circles, and these pixels represent corresponding data records. Visual indicators are assigned to the pixels based on sentiments expressed with respect to a respective topic by the data records represented by the pixels. The pixels are located in regions of the bubbles with different cross hatchings, which represent different colors assigned to the pixels. Pixels in the upper portions of the pixel 200 are assigned color(s) corresponding to a negative sentiment, pixels in the lower portions of the bubble 200 are assigned color(s) corresponding to a positive sentiment, and pixels in a portion between the upper and lower portions are assigned a color corresponding to a neutral sentiment. The pixels are sorted in the bubble 200 according to sentiment scores assigned to the pixels, such that groups of pixels of common color are formed (the groups corresponding to the portions of the bubble 200 of Fig. 2).

[0024] The bubble 200 has various dimensions, including a height (H), an overall width (W1), and a center width (W2). The number of data records represented by the bubble 200 determines the size and shape of the bubble. The area of the bubble 200 is set to allow the bubble 200 to contain the number of pixels representing the data records corresponding to the bubble 200. Also, the overall width W1 of the bubble 200 is based on the time interval represented by the bubble 200.

[0025] In some examples, the height (H) of the bubble 200 is restricted according to the following condition to avoid a bubble having an irregular shape, and to obtain more space along the axis 102 (Fig. 1) of the visualization 100 to depict topics:

$$H \leq \frac{2}{3} W1. \quad (\text{Eq. 1})$$

[0026] Although a specific relationship between H and W1 is expressed in Eq. 1, it is noted that in other examples, a different relationship can be specified between H and W1.

[0027] In some examples, a bubble can be drawn using a B-Spline function. A B-Spline function employs six control points, P1, P2, P3, P4, P5, and P6 shown in Fig. 2. The control points P1 and P2 are separated by the center width (W2). Similarly, the control points P4 and P5, are separated by the center width (W2).

[0028] In other examples, other types of functions can be used for drawing the bubble 200.

[0029] In other examples, instead of having the shape 200 shown in Fig. 2, a bubble can have a different shape, such as a circular shape, an elliptical shape, a polygonal shape, and so forth.

[0030] The ensuing discussion refers to various processes. In some examples, some of the processes can be performed by a client device (*e.g.* client device 602 in Fig. 6), while other processes can be performed by a server device (*e.g.* server device 604 in Fig. 6). Fig. 6 is discussed further below. A client device can refer to an electronic device (*e.g.* desktop computer, notebook computer, tablet computer, smartphone, personal digital assistant, etc.) that includes or is associated with a display device in which a user can view a visualization (*e.g.* 100 in Fig. 1). A server device can refer to a computer or a collection of computers that a client device can communicate with over a network.

[0031] Fig. 3 is a flow diagram of a visualization process at a client device according to some implementations. The visualization process receives (at 302) selected topics identified from data records. The topics are identified based on scoring candidate terms in the data records according to specified metrics. The selected topics can be considered critical topics or important topics, from the perspective of a given enterprise (*e.g.* business concern, educational organization, government agency, individual, etc.). A “candidate term” can refer to certain words (*e.g.* a noun or a combination of nouns) included in a data record, where the noun or combination of nouns can be a candidate topic to be visualized. Scoring of the candidate terms allows scores to be assigned to the candidate terms, such that candidate terms with higher scores can be selected as topics for visualization. Candidate terms with lower scores can be omitted from visualization.

[0032] The specified metrics used for scoring the candidate terms can include some combination of the following: (1) a frequency metric, which represents the frequency of occurrence of data records pertaining to a respective candidate term, (2) a negativity metric, which represents the negativity of sentiment expressed with respect to a candidate term, (3) a context coherence metric, which indicates whether text expressed by multiple data records

pertaining to a candidate term relate to a common topic, and (4) a user-specified metric, which can be any metric specified by a user to affect scoring of a candidate term.

[0033] The frequency metric for a candidate term is computed by determining the number of data records containing the candidate term within a given time interval. The negativity metric for a candidate term can be determined by summing (or otherwise aggregating) sentiment scores (*e.g.* sentiment scores assigned by users, or sentiment scores derived based on opinion words expressed about the candidate term) for the candidate term in data records within a given time interval. Other types of aggregating of sentiment scores can include averaging the sentiment scores, identifying a median of the sentiment scores, identifying a maximum or minimum of the sentiment scores, or any other type of aggregating.

[0034] An example of a user-specified metric can relate to a threshold specified by a user relating to any attribute that may be contained in a data record. For example, the value of the user-specified metric can be set to a first value if the attribute exceeds the threshold, and to a second value if the attribute does not exceed the threshold. As another example, a user-specified metric can be a time-based metric, which can specify that more recent data records are to be weighted higher than less recent data records. Thus, a score computed for a candidate term can be adjusted based on relative recency of a data record in which the candidate term is included. Other examples of user-specified metrics can be used in other implementations. More generally, a user-specified metric can be provided based on user input to allow a user to control scoring of candidate terms. The user input can be received at an electronic device, or can be included in a file, such as a configuration file.

[0035] The following provides a further discussion of the context coherence metric. Assume a set of data records that each mentions a specific candidate term, *e.g.* a “USB cable” term. At this point, it is uncertain whether the data records that mention “USB cable” refer to the same issue, such as a missing USB cable. For example, one subset of the data records can refer to a yellow USB cable, while another subset of the data records can refer to a missing USB cable. Context coherence is determined by checking whether words besides “USB cable” mentioned in the set of data records are common in at least a majority of the data records in the set. As an example, one such word can be “missing,” which can be in close

proximity to the “USB cable” term in some of the data records. The presence of the word “missing” in close proximity to the “USB cable” term in a majority of the data records can indicate with some likelihood that “USB cable” is used in a common context, in other words, relate to the same issue (*e.g.* the issue of a missing USB cable). More specifically, the context coherence metric is based on whether co-occurrence of words in data records (*e.g.* user reviews) relate to the same issue (*i.e.* a respective topic).

[0036] The score (SCORE) that is assigned to a candidate term can be expressed as follows, in some examples:

$$\text{SCORE} = \text{FREQUENCY} \cdot \text{SENTIMENT_NEGATIVITY} \cdot \text{CONTEXT_COHERENCE} \cdot \text{USER_METRIC}, \quad (\text{Eq. 2})$$

where FREQUENCY is the frequency metric, SENTIMENT_NEGATIVITY is the negativity metric, CONTEXT_COHERENCE is the context coherence metric, and USER_METRIC is the user-specified metric. In Eq. 2, the score computed for each candidate term is based on a product of the foregoing metrics. In other examples, a different aggregation of the metrics can be performed, where the different aggregation can include a sum, a weighted sum, or some other aggregation function. Also, although the score in Eq. 2 is calculated on an aggregation of four metrics, it is noted that in other examples, a score can be calculated based on a smaller number of metrics, or a larger number of metrics. For example, the score can be calculated based on two or more of the listed metrics, or other metrics.

[0037] As further shown in Fig. 3, the visualization process generates (at 304) a visualization (*e.g.* 100 in Fig. 1) that includes bubbles representing the respective topics, where the bubbles include pixels representing corresponding data records. The visualization process further assigns (at 306) visual indicators (*e.g.* different colors) to the pixels in each bubble according to sentiments expressed in the corresponding data records for the topic represented by the respective bubble. In other examples, visual indicators are assigned to the pixels according to values of other types of attributes contained in the data records.

[0038] Fig. 4 is a flow diagram of data record processing at a server device according to some implementations. The procedure of Fig. 4 is repeated for each data record that is received. The data record processing includes tokenizing (at 402) the data record into

candidate terms. Tokenizing a data record refers to breaking up a data record into a collection of tokens, where some of the tokens can include candidate terms.

[0039] Tasks 404, 406, 408, 410, 412, 414, and 416 are performed iteratively for each of the candidate terms identified by the tokenizing (402). The data record processing determines (at 404) if the currently considered candidate term is in a management database. A management database is a data structure that stores candidate terms that have been previously processed. If the candidate term is not in the management database, then the candidate term is added (at 406) to the management database. In addition, event detection data associated with a candidate term is also stored, such as in the management database or in another data structure. Examples of event detection data can include any one or some combination of the following: a timestamp (to indicate the last time that a data record containing the candidate term was received), sentiment stores for the candidate term, descriptive terms that describe data records containing the candidate term, and so forth. Note that the event detection data can include information relating to the various metrics used for scoring the candidate terms, such as according to Eq. 2 above.

[0040] If the candidate term is determined (at 404) to already be in the management database, then the event detection data associated with the candidate term is updated (at 410). Next, it is determined (at 412) whether the candidate term is considered to correspond to a critical event (in other words, the candidate term is a critical or important topic that is to be visualized). This can be based on a score assigned to the candidate term, such as a score according to Eq. 2. If the score assigned to the candidate term is greater than some specified threshold, or the score assigned to the candidate term is within the top N ($N > 1$) scores, then the candidate term is identified as a topic to be visualized. If the candidate term does not correspond to a critical event, then the data record processing proceeds to the next candidate term (418). However, if the candidate term is considered (at 412) to correspond to a critical event, then the respective data record is stored (at 414), and the topic is marked (at 416) with a flag for visualization. This flag indicates to a visualization process (such as the visualization process of Fig. 3) that the topic is to be visualized in a visualization (e.g. 100 in Fig. 1).

[0041] As noted above, techniques or mechanisms according to some implementations can be implemented in an arrangement that includes a client device and a server device (*e.g.* client device 602 and server device 604 shown in Fig. 6, discussed further below). In such an arrangement, the data record processing of Fig. 4 can be performed at the server device 604, while the visualization process of Fig. 3 can be performed at the client device 602.

[0042] In other examples, instead of performing the tasks of Figs. 3 and 4 separately on the client device and the server device, the tasks can be performed on a common computer (or a common set of computers).

[0043] Fig. 5 illustrates further processing performed at a client device (*e.g.* 602 in Fig. 6), according to some examples. The client device receives (at 502) topics marked with the flag for visualization (as marked at 416 in Fig. 4). Tasks 504, 506, 508, 510, and 512 are iteratively performed for each topic received. The client device determines (at 504) if the topic is already visualized. If not, a new topic is created (at 506) in the visualization. If the topic is already visualized, then a new topic does not have to be created.

[0044] Next, the client device creates (at 508) a bubble for data records associated with the topic. The bubble is then arranged (at 510) in the visualization. The process then proceeds to the next topic (512).

[0045] As data records are received, the visualization process can update the visualization in respective time intervals (*e.g.* every second, every minute, every hour, etc.). To do so, the timestamp of a first data record is recorded as the beginning of a current time interval. The end of the current time interval can be a time equal to the timestamp corresponding to the beginning of the first interval plus the length of the interval, where the length can be a specified length (*e.g.* set by a user, or preset in a system). Whenever a new data record is received, the system checks whether the data record is still in the current time interval. If the newly received data record is in the current time interval, the data record is processed for visualization in a bubble corresponding to the current time interval. However, if the newly received data record is not in the current time interval, then the visualization is updated to create a new time interval, such that the newly received data record is visualized in a bubble in the new time interval.

[0046] Arranging a bubble in the visualization, as performed at 510 in Fig. 5, can include merging multiple bubbles or splitting a bubble into multiple bubbles. Two bubbles are merged when at least the width of at least one of the bubbles increases such that the two bubbles overlap. In response to detecting overlap of the two bubbles, the two bubbles can be merged into one bubble.

[0047] A given bubble is split into multiple bubbles if there is a time gap in the given bubble with no pixels (in other words, there are no data records relating to a respective topic in the time gap). If such a time gap is detected, then the given bubble is split into two bubbles separated by the time gaps.

[0048] As new data records are received for inclusion in a bubble, the size and shape of the bubble can be updated, by increasing the width (W1) and/or height (H) of the bubble. The size and shape of the bubble is updated in a manner that is consistent with Eq. 1, for example.

[0049] In some examples, the topics in the visualization 100 can be arranged such that topics for which more recent data records have been received are arranged closer to the top of the visualization 100 than topics for which data records have not been recently received. The older topics (those topics for which data records have not been received for some time) are moved closer to the bottom of the visualization 100. However, an older topic that is near the bottom of the visualization 100 can be moved closer to the top of the visualization 100 if data records referring to the older topic are recently received.

[0050] Thus, as data records are received, the positions of the topics (and thus the corresponding bubbles) can continually change. In addition, as data records are received, the size and shape of certain bubbles can also continually change. Thus, the visualization (*e.g.* 100 in Fig. 1) is dynamically updated as data records are received.

[0051] Fig. 6 is a block diagram of an example arrangement that includes the client device 602 and the server device 604, which can be coupled over a network 606. The client device 602 includes a display device 608, which can display a visualization 610 (*e.g.* 100 in Fig. 1). The client device 602 also includes one or multiple processors 612, which can be

coupled to a network interface 614 (for communications over the network 606), and a non-transitory machine-readable or computer-readable storage medium (or storage media) 616. A processor can include a microprocessor, microcontroller, processor module or subsystem, programmable integrated circuit, programmable gate array, or another control or computing device.

[0052] The storage medium (or storage media) 616 can store visualization instructions 618, which are executable on the processor(s) 612 to perform various tasks discussed above including tasks of Figs. 3 and 5, as examples. Moreover, note that in some implementations (such as implementations without the server device 604), the visualization instructions 618 can also perform tasks of Fig. 4.

[0053] The server device 604 includes one or multiple processors 620, which can be coupled to a network interface 622 (for communications over the network 606), and to a non-transitory machine-readable or computer-readable storage medium (or storage media) 624. The storage medium (or storage media) 624 can store data record processing instructions 626 and topic detection instructions 628. The data record processing instructions 626 can perform the processing of Fig. 4, for example. The topic detection instructions 628 can apply the scoring function according to Eq. 2, to assign scores to candidate terms for the purpose of determining topics for visualization based on the scores.

[0054] The storage medium (or storage media) 624 can also store various data records 630, as well as a management database 632. The management database 632 is the management database referred to above for storing candidate terms (and respective event detection data) that have been previously processed.

[0055] The storage media 616 and 624 can include different forms of memory including semiconductor memory devices such as dynamic or static random access memories (DRAMs or SRAMs), erasable and programmable read-only memories (EPROMs), electrically erasable and programmable read-only memories (EEPROMs) and flash memories; magnetic disks such as fixed, floppy and removable disks; other magnetic media including tape; optical media such as compact disks (CDs) or digital video disks (DVDs); or other types of storage devices. Note that the instructions discussed above can be provided on one computer-

readable or machine-readable storage medium, or alternatively, can be provided on multiple computer-readable or machine-readable storage media distributed in a large system having possibly plural nodes. Such computer-readable or machine-readable storage medium or media is (are) considered to be part of an article (or article of manufacture). An article or article of manufacture can refer to any manufactured single component or multiple components. The storage medium or media can be located either in the machine running the machine-readable instructions, or located at a remote site from which machine-readable instructions can be downloaded over a network for execution.

[0056] In the foregoing description, numerous details are set forth to provide an understanding of the subject disclosed herein. However, implementations may be practiced without some of these details. Other implementations may include modifications and variations from the details discussed above. It is intended that the appended claims cover such modifications and variations.

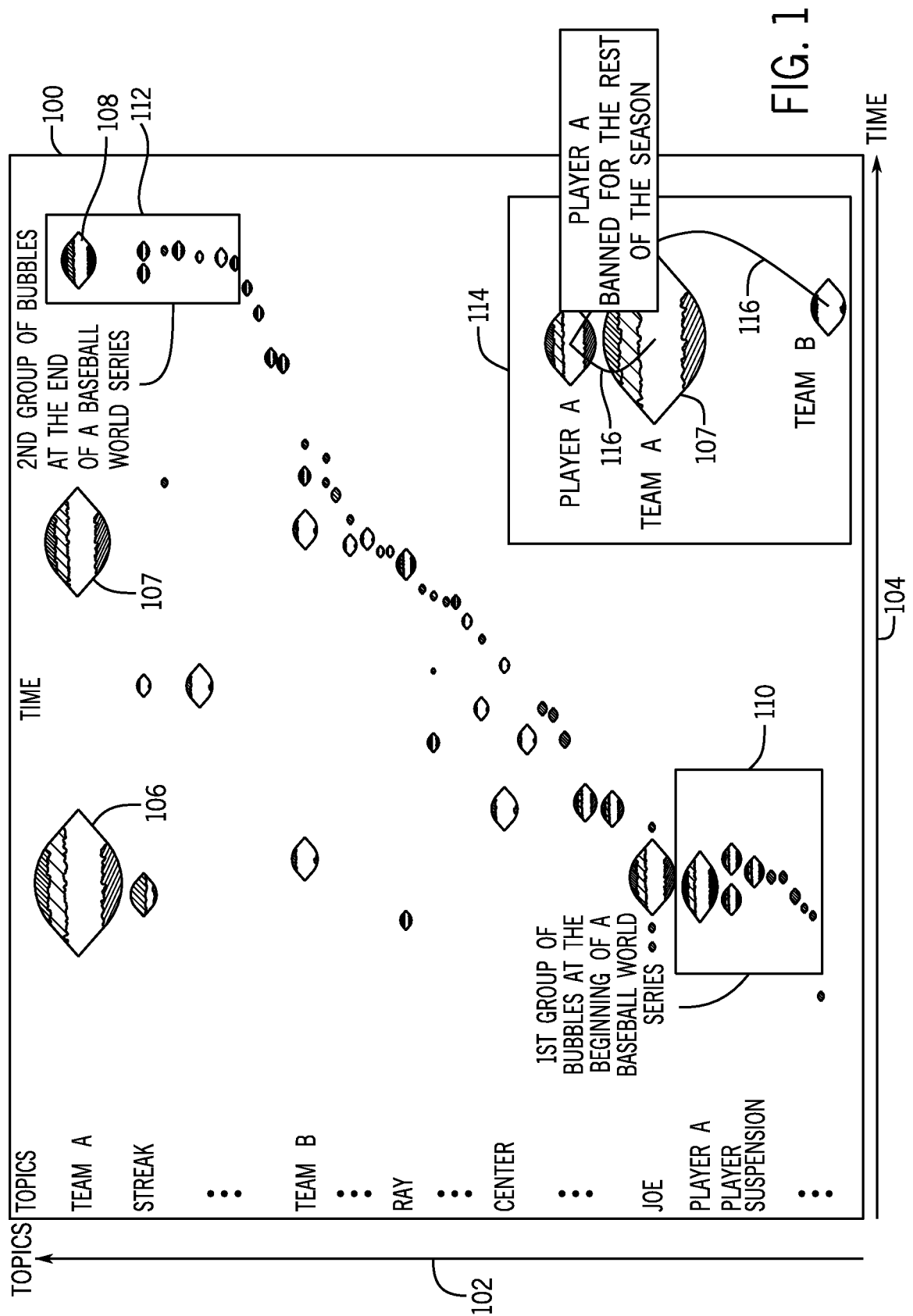
What is claimed is:

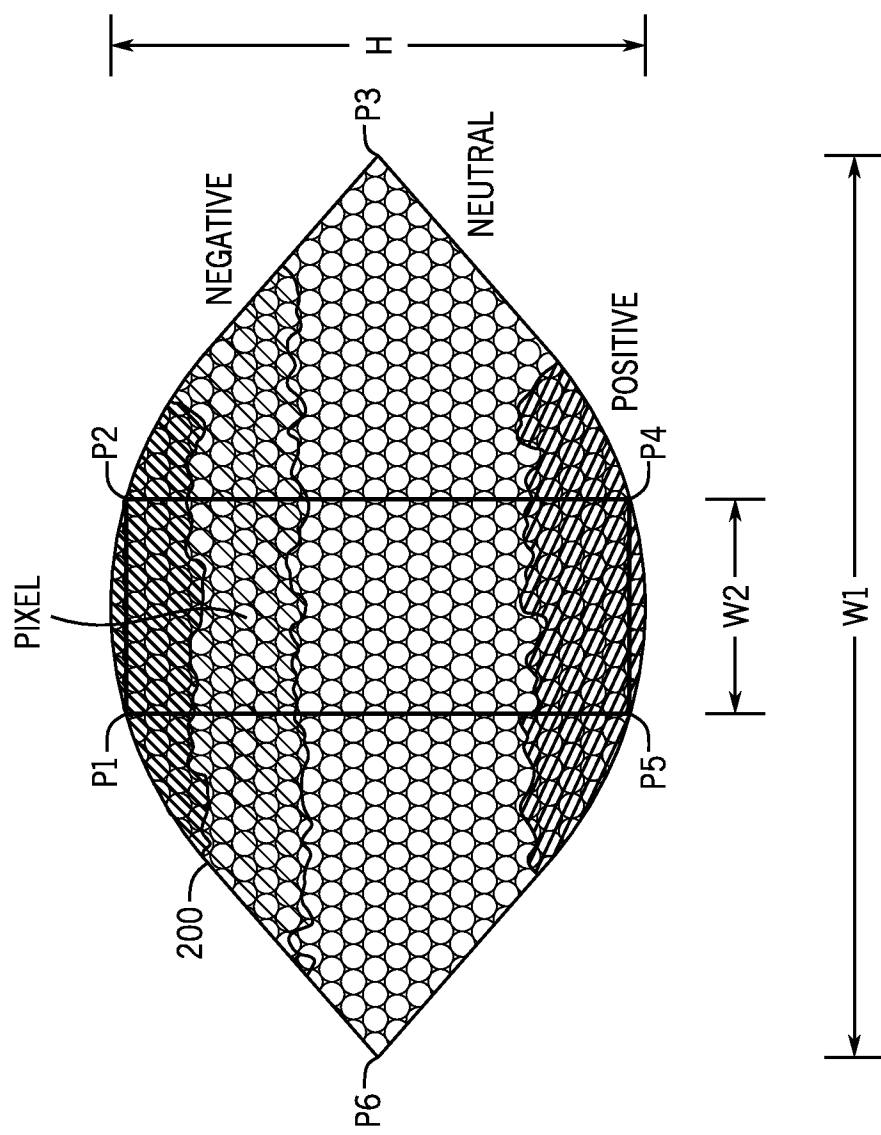
- 1 1. A method comprising:
2 receiving, by a system including a processor, selected topics identified from records
3 based on scoring candidate terms in the records according to a user-specified metric received
4 by the system from a user input, and at least one further metric selected from among
5 frequencies of occurrence of records pertaining to the respective candidate terms, and
6 negativity of sentiment expressed with respect to the candidate terms in the records;
7 generating, by the system, a visualization including bubbles representing the
8 respective topics, the bubbles including pixels representing corresponding records, wherein a
9 given one of the bubbles has a shape dependent upon a number of records represented by the
10 given bubble and a time interval represented by the given bubble; and
11 assigning, by the system, visual indicators to the pixels in the given bubble according
12 to values of an attribute expressed in the corresponding records for the topic represented by
13 the given bubble.
- 1 2. The method of claim 1, wherein the scoring of the candidate terms in the records is
2 further according to a context coherence metric that indicates whether a candidate term
3 included in multiple ones of the records relate to a common topic.
- 1 3. The method of claim 1, wherein the visualization is an interactive visualization, the
2 method further comprising:
3 receiving a user selection of at least one bubble in the interactive visualization;
4 in response to the user selection, generating a zoomed view of the at least one bubble
5 linked with at least another bubble.
- 1 4. The method of claim 1, wherein the generating and the assigning are performed in
2 real time as the records are received.
- 1 5. The method of claim 1, wherein the system includes a client device, and the selected
2 topics are received from a server device that identified the selected topics.

- 1 6. The method of claim 1, wherein a first axis of the visualization corresponds to the
2 identified topics, and a second axis of the visualization corresponds to time intervals, the
3 method further comprising:
4 arranging a plurality of the bubbles along a line parallel to the second axis in the
5 visualization, wherein the plurality of the bubbles relate to one of the identified topics.
- 1 7. The method of claim 1, wherein assigning the different visual indicators comprises
2 assigning different colors according to different sentiments including a positive sentiment and
3 a negative sentiment.
- 1 8. The method of claim 1, wherein the topics are identified by selecting a subset of the
2 candidate terms based on scores produced by the scoring.
- 1 9. The method of claim 1, further comprising merging at least two of the bubbles in
2 response to at least one of the at least two bubbles growing in size such that the at least two
3 bubbles overlap.
- 1 10. The method of claim 1, further comprising splitting a particular one of the bubbles
2 into multiple bubbles in response to detecting that a time gap with no records is present in the
3 particular bubble.

- 1 11. A system comprising:
2 at least one processor to:
3 receive records including user feedback;
4 identify candidate terms in the records;
5 score the candidate terms according to a user-specified metric received by the
6 system from a user input, and at least one further metric selected from among frequencies of
7 occurrence of records pertaining to the respective candidate terms, and negativity of
8 sentiment expressed with respect to the candidate terms in the records;
9 select, based on the scoring, a subset of the candidate terms as topics; and
10 generate a visualization including bubbles representing the respective topics,
11 the bubbles including pixels representing corresponding records, wherein a given one of the
12 bubbles has a size and shape dependent upon a number of records represented by the given
13 bubble and a time interval represented by the given bubble, and wherein the pixels in the
14 given bubble are assigned visual indicators according to sentiments expressed in the
15 corresponding records for the topic represented by the given bubble.
- 1 12. The system of claim 11, wherein the visual indicators include different colors that
2 correspond to a positive sentiment, a neutral sentiment, and a negative sentiment.
- 1 13. The system of claim 12, wherein the at least one processor is to sort the pixels within
2 the given bubble by values of a sentiment attribute of the records represented by the given
3 bubble, the sorting causing formation of groups of pixels, each of the groups including pixels
4 of a same respective color.
- 1 14. The system of claim 11, wherein points along a first axis of the visualization
2 correspond to different ones of the topics, and points along a second axis of the visualization
3 corresponds to different times.

1 15. An article comprising at least one non-transitory machine-readable storage medium
2 storing instructions that upon execution cause a system to:
3 receive selected topics identified in records based on scoring candidate terms in the
4 records according to a user-specified metric and at least two metrics selected from the group
5 consisting of: frequencies of occurrence of records pertaining to the respective candidate
6 terms, negativity of sentiment expressed with respect to the candidate terms in the records,
7 and a context coherence metric that indicates whether text expressed by multiple ones of the
8 records pertaining to each respective candidate term relate to a common topic;
9 generate a visualization including bubbles representing the respective identified
10 topics, the bubbles including pixels representing corresponding records, wherein a given one
11 of the bubbles has a size and shape dependent upon a number of records represented by the
12 given bubble and a time interval represented by the given bubble, wherein points along a first
13 axis of the visualization correspond to different ones of the topics, and points along a second
14 axis of the visualization corresponds to different times, and wherein multiple ones of the
15 bubbles are arranged along a line parallel to the second axis, the multiple bubbles
16 corresponding to one of the topics; and
17 assign visual indicators to the pixels in the given bubble according to sentiments
18 expressed in the corresponding records for the topic represented by the given bubble.





3 / 6

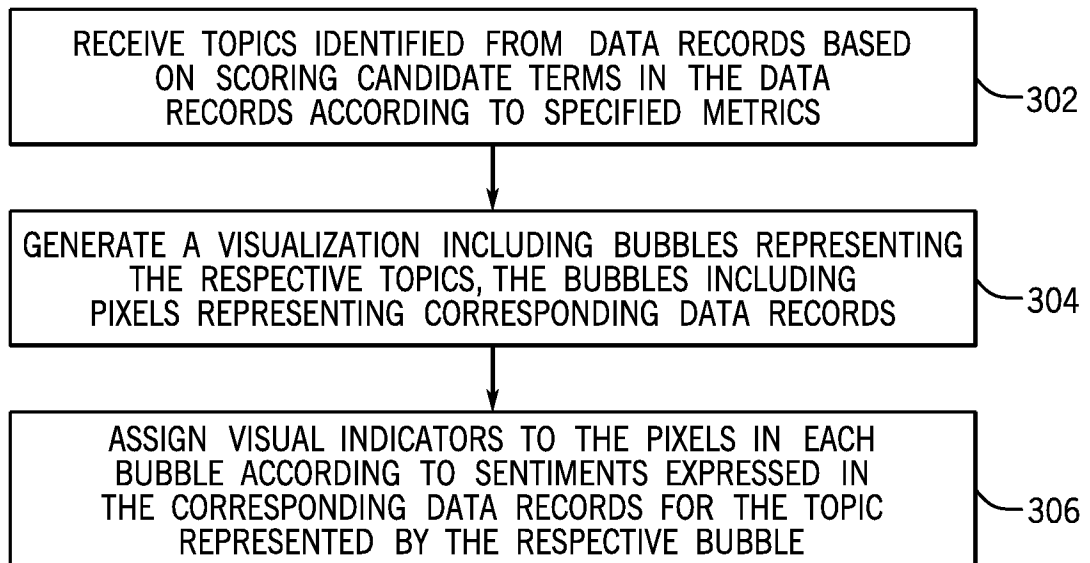


FIG. 3

4 / 6

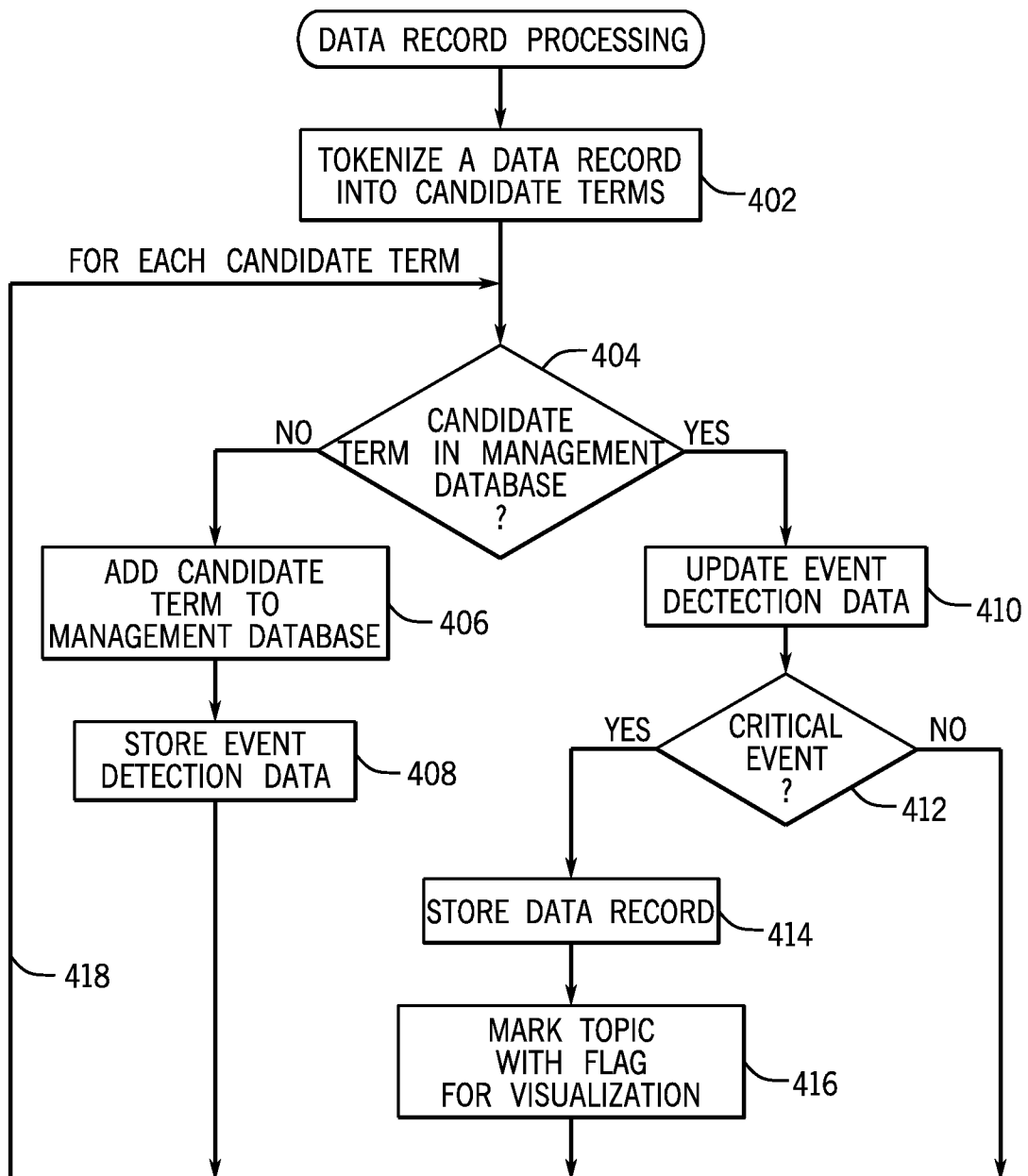


FIG. 4

5 / 6

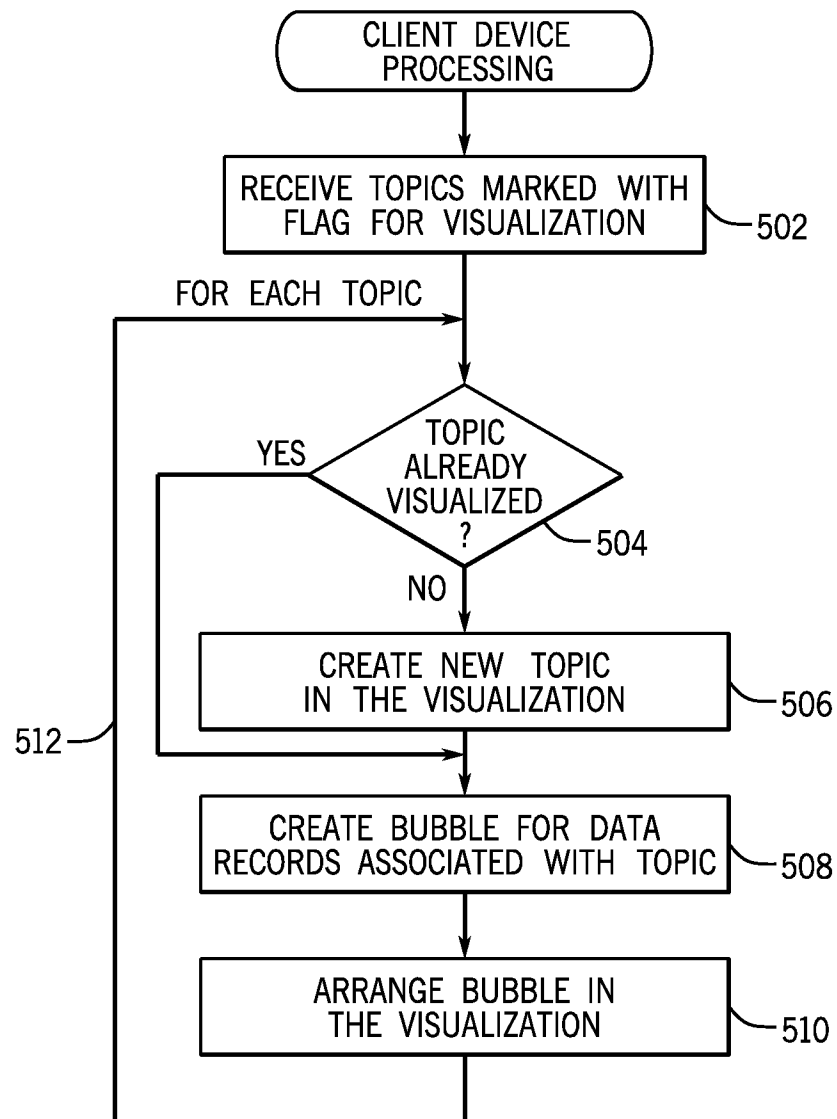


FIG. 5

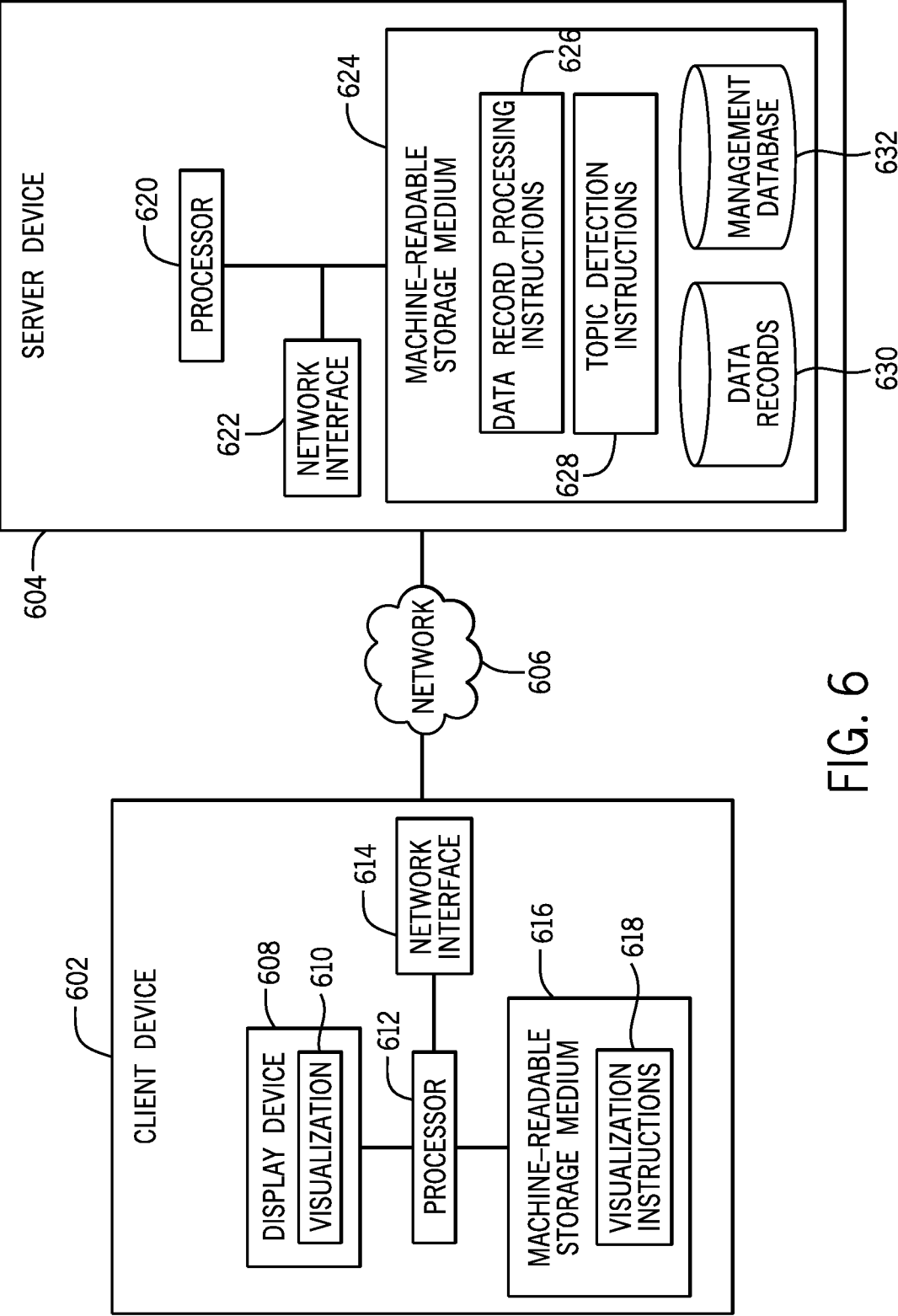


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/036056**A. CLASSIFICATION OF SUBJECT MATTER****G06F 17/00(2006.01)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 17/00; G06F 3/048; G06F 17/30; G06Q 10/00

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: visualization, topic, sentiment, candidate, term, bubble, and similar terms

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2012-0317049 A1 (MING C. HAO et al.) 13 December 2012 See paragraphs [0012]-[0016] and [0026]-[0027]; claims 1, 4-5, 7, 9, 12, and 19; and figure 1.	1-15
A	US 2010-0106752 A1 (RALPH W. ECKARDT, III et al.) 29 April 2010 See paragraph [0017] and [00142]-[00176]; and figures 3 and 8-12.	1-15
A	US 2010-0332465 A1 (JANSSENS FRIZO et al.) 30 December 2010 See paragraphs [0006] and [0092]-[0126]; and figures 6-7.	1-15
A	WO 2012-167399 A1 (HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P.) 13 December 2012 See paragraphs [0007]-[0012] and figure 1.	1-15
A	US 2012-0323627 A1 (RUSSELL ALLEN HERRING, JR. et al.) 20 December 2012 See paragraphs [0004] and [0023]-[0032]; and figure 1.	1-15



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

"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

29 January 2015 (29.01.2015)

Date of mailing of the international search report

29 January 2015 (29.01.2015)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsu-ro, Seo-gu, Daejeon Metropolitan City, 302-701,
Republic of Korea

Facsimile No. ++82 42 472 3473

Authorized officer

NHO, Ji Myong

Telephone No. +82-42-481-8528



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2014/036056

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2012-0317049 A1	13/12/2012	US 8595151 B2	26/11/2013
US 2010-0106752 A1	29/04/2010	CN 101084496 A	05/12/2007
		CN 101084496 B	21/11/2012
		EP 1759280 A2	07/03/2007
		EP 1759280 A4	26/08/2009
		EP 2487599 A1	15/08/2012
		EP 2487600 A1	15/08/2012
		EP 2487601 A1	15/08/2012
		HK 1115720 A1	21/06/2013
		KR 10-1126028 B1	12/07/2012
		KR 10-2007-0085109 A	27/08/2007
		US 2006-0106847 A1	18/05/2006
		US 7672950 B2	02/03/2010
		WO 2005-107405 A2	17/11/2005
		WO 2005-107405 A3	23/08/2007
US 2010-0332465 A1	30/12/2010	CA 2747214 A1	24/06/2010
		EP 2365894 A1	21/09/2011
		EP 2365894 B1	26/06/2013
		EP 2377052 A1	19/10/2011
		EP 2379735 A1	26/10/2011
		JP 2012-512036 A	31/05/2012
		JP 5592391 B2	17/09/2014
		US 2011-0260386 A1	27/10/2011
		US 2012-0052532 A1	01/03/2012
		WO 2010-069047 A1	24/06/2010
		WO 2010-069550 A1	24/06/2010
		WO 2010-078925 A1	15/07/2010
WO 2012-167399 A1	13/12/2012	US 2014-0344243 A1	20/11/2014
US 2012-0323627 A1	20/12/2012	None	