



(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/CN2011/000957

(22) International Filing Date:
8 June 2011 (08.06.2011)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **HEW-LETT-PACKARD DEVELOPMENT COMPANY, L.P.** [US/US]; 11445 Compaq Center Drive W., Houston, TX 77070 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **HAO, Ming, C.** [US/US]; 1501 Page Mill Rd., Palo Alto, California 94304-1100 (US). **DAYAL, Umeshwar** [US/US]; 1501 Page Mill Rd., Palo Alto, California 94304-1100 (US). **ZHOU, Baoyao** [CN/CN]; Room 202, Building 3, No.4 Xibianmenwai Street, Xicheng District, Beijing 100045 (CN). **CHANG, Cheng** [CN/CN]; Jin Quan Jia Yuan, Building No.2, 2-1001, DaTun Road, Chaoyang District, Beijing 100101 (CN). **HSU, Meichun** [US/US]; 1501 Page Mill Rd., Palo Alto, California 94304-1100 (US). **DEKHIL, Mohamed, E.** [US/US]; 1501 Page Mill Rd., Palo Alto, California 94304-1100 (US). **GHOSH, Riddhi-man** [IN/US]; 1501 Page Mill Rd., Palo Alto, California 94304-1100 (US). **ROHRDANTZ, Christian** [DE/DE];

Universitat Konstanz, Fachbereich Informatik und Informationswissenschaft, Box 78, 78457, Konstanz (DE).

(74) Agent: **CHINA PATENT AGENT (H.K.) LTD.**; 22/F, Great Eagle Centre, 23 Harbour Road, Wanchai, Hong Kong (CN).

(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, 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, IS, JP, KE, KG, KM, 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, PE, PG, PH, PL, PT, RO, RS, RU, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: SENTIMENT TREND VISUALIZATION RELATING TO AN EVENT OCCURRING IN A PARTICULAR GEOGRAPHIC REGION

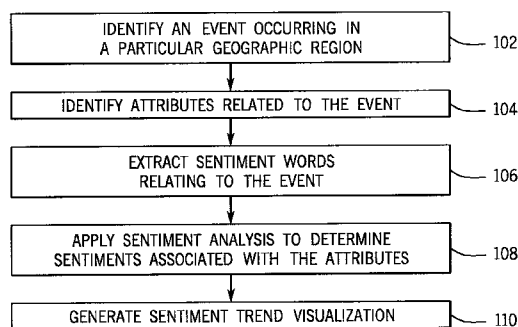


FIG. 1

(57) Abstract: An event occurring in a particular geographic region is identified based on disseminated information containing public commentary in the particular geographic region. Attributes that are related to the event are identified, and sentiment words relating to the identified event are extracted from the disseminated information, where the extracted sentiment words are in a local language of the particular geographic region. A sentiment trend visualization is generated that depicts a trend of sentiments of at least a particular one of the identified attributes, wherein the sentiments are based on the sentiment words for at least the particular attribute.



Sentiment Trend Visualization Relating to an Event Occurring in a Particular
Geographic Region

Background

[0001] Various information can be disseminated with respect to an event, where examples of such disseminated information include news articles (such as articles published by news websites or other news outlets), blogs, online reviews (e.g. customer reviews of products or services), or other types of information. Depending upon the severity of the event or interest of people in the event, there can be a relatively large amount of disseminated information relating to the event.

Brief Description Of The Drawings

[0002] Some embodiments are described with respect to the following figures:

Figure 1 is a flow diagram of a sentiment trend visualization process according to some implementations;

Figures 2, 3A, and 3B depict graphical representations of visualizations according to various implementations;

Figure 4 is a flow diagram of a sentiment trend visualization process according to alternative implementations; and

Figure 5 is a block diagram of the example system incorporating some implementations.

Detailed Description

[0003] For a given event that occurs in a particular geographic region (referred to as “local geographic region” in the ensuing discussion), there can be a relatively large amount of public information disseminated (*e.g.* published or otherwise made available to people who have an interest) that may relate to the event. Generally, an “event” refers to an activity, problem, failure, or any other issue. For example, an enterprise (*e.g.* company, educational organization, government agency, etc.) may have an offering (a product or a service) in a particular market, such as a market within a specific country or region of a country. Due to an event (*e.g.* problem) associated with the offering, various sources may disseminate information regarding public commentary about such problem. “Public commentary” refers to any expression, such as a review, feedback, news article, blog, and so forth expressed in public forum(s).

[0004] News sites may publish news articles regarding the problem associated with the offering. There can also be blogs that discuss the problem associated with the offering. Customer review websites (websites that allow users to enter their feedback about specific products or services) can also provide reviews regarding the offering. Other information disseminators can include social networking sites such as FacebookTM and TwitterTM.

[0005] With the relatively large volume of disseminated information containing public commentary about an event, it can be time-consuming and difficult for users (*e.g.* sales or marketing personnel, analysts, etc.) to manually read through the disseminated information to understand sentiments of individuals with respect to the

event, such as a problem associated with an offering of an enterprise. In fact, disseminated information relating to an event can be in the local language of the local geographic region in which the event occurred. Alternatively, the disseminated information relating to the event can be in multiple languages, such as the local language and another language (e.g. English).

[0006] The sentiments associated with an event may include positive sentiments, negative sentiments, or neutral sentiments. A “sentiment” refers to an attitude, opinion, or judgment of a human with respect to a specific event.

[0007] In accordance with some implementations, sentiment visualization mechanisms or techniques are providing to allow for automated analysis and visualization of multiple sources of disseminated information containing public commentary about one or multiple events occurring in a local geographic region. The sentiment visualization assists users in better understanding sentiments associated with the one or multiple events. In accordance with some implementations, the sentiment visualization is focused on local geographic regions, such as specific countries, states, provinces, cities, and so forth. Thus, rather than processing sources of information regarding particular event(s) disseminated across the world, the sentiment visualization mechanisms or techniques according to some implementations are able to focus on specific local geographic regions of interest.

[0008] Figure 1 is a flow diagram of a process according to some implementations of performing sentiment analysis in a local geographic region. The process includes identifying (at 102) an event occurring in the local geographic region, where the identifying of the event is based on disseminated information

containing public commentary in the local geographic region. For example, the disseminated information can be from local news outlets (local news websites, local newspapers, local television or radio stations, etc.). The disseminated information can also be from local blogs or local social networking posts, where a local blog or a local social networking post refers to a blog or post that focuses on events of the local geographic region. The disseminated information can be received in the form of data records, where each data record contains respective commentary provided by the different sources of information, *e.g.* news outlets, blogs, social networking posts, etc.

[0009] The identified event can be an event that was previously not known to an analyst (a person who is interested in understanding user sentiment). For example, an enterprise may be selling a product in a local geographic region. The product may exhibit a problem that is being experienced by a large number of users (but this problem is not yet known to the enterprise). As information about this problem becomes disseminated, the problem can be discovered and identified. Thus, according to some implementations, not only can past events be identified, but potential events can also be identified for sentiment analysis. A potential event can be an event that is not known by an analyst (or enterprise) to exist, but is likely to happen or is in the process of occurring.

[0010] The process of Figure 1 further identifies (at 104) attributes that are related to the event. The attributes are found in the disseminated information (in the received data records), and the identified attributes are those attributes satisfying predefined one or multiple criteria. An attribute can be a noun or compound noun (a

noun formed of multiple words, such as “customer service”) that exists in the disseminated information. Alternatively, an attribute can be derived from nouns or compound nouns in the disseminated information. Another criterion for selecting an attribute can be that the attribute should be associated with sentiment words, where a “sentiment word” refers to any word or phrase found in the disseminated information that expresses sentiments of individuals. Elements in the disseminated information that are not associated with sentiment words can be disregarded as not being useful attributes.

[0011] Sentiment words in the local language are extracted (at 106) from the disseminated information. Sentiment words include individual words or phrases (made up of multiple words) that express an attitude, opinion, or judgment of a human. Examples of sentiment words include “bad,” “poor,” “great performance,” “fast service,” and so forth. Based on the sentiment words and the identified attributes, sentiment analysis can be applied (at 108) to determine the sentiments associated with the attributes in the local geographic region. The sentiment analysis can use any one of various sentiment analyzers that are generally available.

[0012] A sentiment trend visualization is then generated (at 110) that depicts a trend of the determined sentiments of at least one of the identified attributes. The sentiment trend visualization can also depict multiple ones of the identified attributes.

[0013] Figure 2 is a graphical representation of an example sentiment trend visualization 200 that can be generated according to task 110 in Figure 1. The sentiment trend visualization 200 includes various display areas, including an attribute list area 202 that contains a list of attributes that can be ordered according

to some importance factor. The ordering allows the more important attributes to be provided at the upper portion of the attribute list, while the less important attributes are lowered down in the attribute list. To view the less important attributes, a user can use a scroll bar 204 (shown on the right of the attribute list) to scroll through the attribute list in the attribute list area 202.

[0014] A trend graph area 206 is also part of the sentiment trend visualization 200. In the example of Figure 2, a first graph 208 depicts positive sentiments associated with a particular attribute, while a second graph 210 depicts negative sentiments associated with the same particular attribute. In the example of Figure 2, the particular attribute associated with the graphs 208 and 210 is the highlighted attribute 212 ("notebook" attribute) in the attribute list area 202.

[0015] The horizontal axis of each graph 208, 210 depicted in the sentiment trend graph area 206 represents time, whereas the vertical axis of each graph represents an aggregate sentiment score. In the example of Figure 2, vertical dividing lines are shown in the trend graph area 206 to divide the total time period represented by graphs 208, 210 into discrete months (January through May 2010 in the example of Figure 2).

[0016] Each point on the graph 208 or 210 can represent a respective aggregate score, derived by aggregating (e.g. summing or performing some other aggregate, such as averaging, identifying a median, etc.) sentiment scores associated with different data records within some predefined time interval, such as one hour, one day, one week, etc. Thus, the points along each graph 208 or 210 can be thought of as points in successive time intervals, where within each time interval, sentiment

scores for the respective attribute (where each sentiment score is computed based on content of the data records) are aggregated to produce an aggregate score that is plotted on the graph. For example, point 212 of the positive graph 208 indicates a high positive aggregate sentiment score within a time interval (e.g. one hour, one day, etc.) in January 2010.

[0017] A user can quickly ascertain that the sentiment scores associated with the attribute being depicted in the trend graph area 206 are generally more positive than negative (the positive aggregate scores on the positive graph 208 are generally larger than the negative aggregate scores on the negative graph 210). Also, the graphs 208 and 210 indicate that stronger sentiments were expressed earlier in 2010 (January through first part of April), with sentiments being more neutral later in 2010 (in late April and in May). The graphs 208 and 210 have generally larger aggregate sentiment scores during the early part of 2010, and generally lower aggregate sentiment scores in the time frame depicted generally as 214 in Figure 2 (which covers the latter part of April and May).

[0018] The time period displayed in the trend graph area 206 is selectable. A graph 222 below the trend graph area 206 extends over a larger time period—a user can select a portion of the graph 222 (such as portion 220) for further details to be displayed in the trend graph area 206. The graph 222 can be either a positive graph or negative graph.

[0019] The sentiment trend visualization 200 further includes a text display area 230, which can display actual text of commentary in one or multiple data records relating to the selected attribute 212 (“notebook” attribute).

[0020] Figure 3A depicts a different example sentiment trend visualization 300 that can be generated using trend visualization mechanisms or techniques according to some implementations. In Figure 3A, various vertical bars are shown. Each vertical bar 302 corresponds to a respective attribute 304. The visualization 300 of Figure 3A shows a trend over multiple attributes, whereas the visualization 200 of Figure 2 shows a trend over time.

[0021] The horizontal axis of the visualization 300 represents different attributes, whereas the vertical axis represents numbers of positive and negative data records. As noted above, each data record represents a respective piece of disseminated information (*e.g.* news article, blog entry, social networking post, etc.). For a given attribute, a positive data record refers to a data record that expresses a positive sentiment regarding the given attribute, while a negative data record refers to a data record that expresses a negative sentiment regarding the given attribute. Note that a data record can contain sentiment words that express positive and negative sentiments regarding different attributes.

[0022] Each vertical bar 302 includes an arrangement of cells, where each cell represents a respective data record. Each cell is assigned a respective visual indicator, which in Figure 3A is a respective color. The visual indicator (*e.g.* color) assigned to each cell depends upon the corresponding sentiment score associated with the respective attribute. A color scale 306 in Figure 3A indicates what the different colors represent. According to the color scale 306, the green colors correspond to positive sentiment scores, whereas the pink/red colors correspond to negative sentiment scores. The gray color corresponds to a neutral sentiment score.

[0023] Within each bar 302, the cells associated with negative sentiment scores are provided in a first portion (e.g. lower portion of the bar), whereas the cells associated with positive sentiment scores are provided in a second portion (e.g. upper portion of the bar). The cells associated with neutral sentiment scores can be provided between the first and second portions. In alternative examples, the green cells can be provided at the lower portion of the bar, while the pink/red cells can be provided at the upper portion of the bar. Generally, the cells in each bar 302 are ordered according to respective sentiment scores associated with corresponding data records—such ordering allows the cells having similar or the same sentiment scores to be grouped together into the green portion, gray portion, and pink/red portion of each bar 302.

[0024] The vertical axis of Figure 3A includes positive numbers and negative numbers. The positive numbers along the vertical axis indicates numbers of data records (positive data records) that have expressed positive sentiments for a given attribute, while the negative numbers along the vertical axis indicates numbers of data records (negative data records) that have expressed negative sentiments for a given attribute. In the example of Figure 3A, for the “HP” attribute, there are about 400 positive data records and about 170 negative data records.

[0025] The length of each bar 302 represents the total number of data records. A number is provided near the top of each bar 302 (a row 320 of numbers is shown in Figure 3A). For the bar 302 corresponding to the “HP” attribute, the number “1017” indicates that the bar 302 contains cells representing 1017 data records. The bar 302 for to the “user” attribute includes cells corresponding to 567 data records,

and the bar for the “product” attribute includes cells corresponding to 471 data records. The remaining bars 302 in the visualization 300 are also associated with respective total data record numbers.

[0026] Using the bar chart shown in the trend visualization 300 of Figure 3A, a user can quickly determine which attribute is the one associated with the largest numbers of data records (*i.e.* largest number of sources of disseminated information). Also, for each attribute, a user can quickly determine whether the different sources of information are largely positive or negative or neutral.

[0027] The content of each cell can be depicted by moving a cursor over the respective cell. For example, arrow 310 in Figure 3A indicates that a cursor has been moved over a particular cell, and the content of the data record corresponding to that cell is represented in a pop-up table 312. The pop-up table includes various fields and associated values for those fields. In the example of Figure 3A, the pop-up table 312 is for a cell in the bar 302 for the “notebook” attribute.

[0028] Since the sentiment visualization according to some implementations is for a local geographic region, it is possible that the attributes found in the data records can be in a language other than English. The sentiment visualization techniques or mechanisms according to some implementations are able to provide visualization of attributes in a non-English language.

[0029] For example, Figure 3B shows the same trend visualization 300 as in Figure 3A, except that the attributes 304 are in the Chinese language (assuming that the local geographic region of interest is a Chinese-speaking country, such as China). The pop-up table 312 also contains some values that are in the Chinese language.

[0030] Figure 4 is a flow diagram of a process according to further implementations. Local disseminated information 402 (in the form of data records) relating to a local geographic region is available in a database (or multiple databases). Also, the local disseminated information 402 can be continually received for analysis, such as from websites, social networking sites, and so forth.

[0031] From the local disseminated information 402, local event selection (404) is performed. Local event selection involves identifying at least one event based on the content of the local disseminated information 402. The identified at least one event can include known event(s) and/or potential event(s) (event(s) not yet known but are likely to occur or are in the process of occurring). Texts relating to the selected at least one event are stored (406).

[0032] Next, from the local disseminated information 402, local sentence detection (408) is performed to identify sentences in the local disseminated information. The identified local sentences (in the local language) are stored (410).

[0033] The local disseminated information 402 is also processed to perform local tokenization (412). Tokenization involves breaking an input collection of information (in this case the local disseminated information 402) into meaningful parts, such as words. In some implementations, the tokenization (412) employs a local dictionary (or local dictionaries) to assist in identifying words that can potentially constitute or define attributes that relate to the identified event(s). The local words derived from the tokenization are stored (414).

[0034] Next, local part-of-speech (POS) tagging (416) is performed on the identified local words. The POS tagging marks up (tags) words identified by the

tokenization (412) with a corresponding part of speech (e.g. noun, verb, adverb, adjective, etc.). POS tagging can be based on a definition of the word (as extracted from a local dictionary) as well as a context of the word (e.g. relationship of the word with adjacent words in a phrase, sentence, or paragraph). After the POS tagging (416), the words with local POS tagging are stored (418).

[0035] Based on the POS-tagged words (stored at 418), the process of Figure 4 can perform various processes, including detection of local compound phrases (420), detection of local negation (422), and detection of local sentiment words (424). The detection of local compound phrases (420) identifies nouns and noun phrases that potentially can be attributes that are related to the selected local event(s). Examples of identified nouns and noun phrases (stored at 426) include “notebook,” “customer service,” “problem,” and so forth. These nouns and noun phrases can be attributes on which users have expressed sentiments that relate to the selected local event(s).

[0036] The detection of local negation (422) produces words that express negative sentiments (e.g. bad, poor, fast, etc.). Note that the word “fast” can express either a positive sentiment or negative sentiment, depending on the context. If “fast” is used in the context of describing customer service or product performance, then “fast” expresses a positive sentiment. On the other hand, if “fast” is used in the context of describing usage of an ink cartridge, then “fast” expresses a negative sentiment. Thus, the detection of local negation (422) is based on determining the appropriate context in which any sentiment word is expressed. The negation words (words that express a negative sentiment) are stored (428).

[0037] The detection of local sentiment words (424) identifies words or phrases including multiple words that express sentiments (whether positive, negative, or neutral). These words or phrases potentially can express sentiments regarding various attributes that relate to the selected event(s).

[0038] From the local nouns and/or noun phrases (stored at 426), the process of Figure 4 performs attribute extraction (432) based on the selected event(s). Attribute extraction involves selecting a set of attributes from a larger group of candidate attributes. The selected attributes are stored (434). The stored selected attributes can be associated with corresponding sentiment words (from 430), where each sentiment word can be an individual word or a phrase.

[0039] The selection of attributes can involve selection of attributes that are considered to be the most important to the selected event(s). In some implementations, the selection of attributes can be based on various selection criteria, including, as examples, at least some combination of the following: a criterion relating to how frequently each attribute occurs in the input data records, a criterion relating to relative amounts of negative and positive feedback associated with each attribute, a criterion relating to time density of feedback associated with each attribute, a criterion relating to negativity associated with each attribute, and an application-specific criterion. "Feedback" refers to commentary provided about an attribute in each of the data records containing the local disseminated information 402.

[0040] The selection criterion for selecting an attribute from among multiple attributes relating to frequency of occurrence is referred to as a "high-frequency

criterion". The frequency of occurrence of an attribute refers to how frequently the attribute occurs in the input data records. An attribute is considered to occur frequently in the input data records if an amount of occurrence of the attribute in the data records exceeds some predefined threshold. For example, the amount of occurrence of the attribute can be expressed as a percentage of the data records in which the attribute is present—thus, the attribute can be considered to be a high-frequency attribute if the percentage of data records in which the attribute appears is greater than some predefined percentage threshold. A high-frequency attribute is assigned a greater importance (e.g. higher importance score).

[0041] As listed above, another selection criterion considered in selecting attributes is relative amounts of negative and positive feedback associated with each attribute (referred to as a "large difference criterion"). A given attribute can have a first number of positive feedback and a second number of negative feedback. The number of positive or negative feedback can refer to the number of data records in which the attribute appears (such as within a particular time interval being studied) and for which users have expressed positive or negative feedback, respectively. The number of positive or negative feedback can also refer to the number of times that users have expressed positive or negative review, respectively, of the given attribute in the input data records. An attribute associated with a large difference between the number of positive feedback and the number of negative feedback is assigned greater importance (e.g. higher importance score).

[0042] As noted above, another selection criterion considered in selecting a subset of attributes from among candidate attributes is the time density of feedback

associated with each attribute (referred to as a “time density criterion”). “Time density” refers to the average time gap between successive data records that contain feedback for a given attribute. For example, the average time gap between feedback for a first attribute may be X , while the average time gap between feedback data records for a second attribute is Y , where $X < Y$. In such an example, the first attribute is considered more important than the second attribute, since the average time gap between reviews for the first attribute is smaller than the average time gap for the second attribute. This is an indication that users are more interested in the first attribute (and thus have expressed more feedback on the first attribute). Alternatively, techniques can consider the change in time density—a sudden increase in time density of an attribute indicates that the attribute would be more important.

[0043] Another selection criterion is a criterion relating to negativity associated with each attribute. An attribute associated with greater negativity is assigned greater importance than another attribute with less negativity. The “negativity” of an attribute refers to the amount of negative feedback received for the given attribute. For example, within 1,000 data records, a first attribute may be associated with 500 negative reviews, whereas a second attribute may be associated with 100 negative reviews. In this case, this selection criterion (referred to as the “negativity criterion”) would indicate that the first attribute would be more important than the second attribute. Alternatively, instead of a negativity criterion, a positive criterion can be more relevant in certain application—an attribute with a greater number of positive reviews would be more important.

[0044] Another selection criterion that can be considered for selecting a subset of attributes from among candidate attributes is the application-specific criterion. An “application-specific criterion” refers to a criterion indicating that, for a given application, certain attributes are more important than other attributes. For example, if the application is sentiment analysis for a notebook computer, then the application-specific criterion can specify a certain group of attributes that should be considered for analysis, such as price, CPU performance, display size, etc.

[0045] The multiple selection criteria (high-frequency criterion, large difference criterion, time density criterion, negativity criterion, and application-specific criterion) can be combined to assign an importance score to each candidate attribute. Each of the selection criteria causes an individual score to be produced for the candidate attribute. Such individual scores can be combined (such as by use of a weighted sum) to produce an overall score indicating the relative importance of the candidate attribute in user sentiment understanding. A weighted sum approach involves assigning respective weights to the individual scores (the weights are multiplied to the individual scores, and the weighted individual scores are summed to produce the overall score). The weights assigned can be equal weights, or alternatively, the weights assigned can be different.

[0046] Once attributes have been selected (and sentiment words have been associated with the selected attributes) and stored (at 434), local sentiment analysis (436) can be performed. Various available sentiment analyzers can be used to perform the sentiment analysis, given the selected attributes and associated sentiment words/phrases (434).

[0047] The local sentiment analysis (436) produces attributes with sentiment scores (stored at 438). As a result of the sentiment analysis (436), each of the selected attributes (selected at 432) is assigned a respective sentiment score.

[0048] The sentiment scores of an attribute can vary over time, as shown in Figure 2. At a first point in time, users may provide more intense feedback regarding a given attribute (where “more intense” feedback refers to more positive or negative feedback associated with a sentiment score having a greater magnitude). Later, such as in response to a problem being resolved, the feedback becomes less intense.

[0049] In some cases, sentiment scores for each attribute can be aggregated (440) for each respective time interval (e.g. hour, day, week, etc.). For example, an attribute score for each attribute can be computed for each hour. However, to provide more meaningful visualization (or less clustered visualization), the sentiment scores computed in successive hours can be aggregated into an aggregate score for a given day (made up of 24 hours). Such aggregate scores are depicted in a trend visualization, such as the trend visualization 200 of Figure 2.

[0050] The process of Figure 4 next performs visual analysis (442) to produce one or multiple trend visualizations of the attributes and their associated sentiment scores. The visualizations include bar chart visualizations 444 (such as those shown in Figures 3A and 3B) and sentiment trend visualizations 446 (such as shown in Figure 2).

[0051] By employing techniques and mechanisms according to some implementations, a sentiment trend of specific attributes relating to selected event(s)

can be automatically discovered and can be visualized for convenient analysis, without having to manually read through potentially large amounts of disseminated information, which can be in a non-English language.

[0052] Figure 5 shows an example computer system 500 (which can be implemented as an individual computer or as a distributed system of computers) that has a visual analytics module 502, which is capable of performing the tasks of Figures 1 and 4, for example. The visual analytics module 502 is executable on one or multiple processors 504, which is (are) connected to storage media 508 and to a network interface 506.

[0053] The storage media 508 store various information 510, such as those discussed above in connection with Figure 4. The network interface 506 allows the computer system 500 to communicate over a data network 512 with remote systems 514. The remote systems 514 can be sources of disseminated information that can be retrieved by the computer system 500.

[0054] A processor can include a microprocessor, microcontroller, processor module or subsystem, programmable integrated circuit, programmable gate array, or another control or computing device.

[0055] The storage media 508 are implemented as one or multiple computer-readable or machine-readable storage media. The storage media 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 or all 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 of a system having a processor, comprising:
2 identifying an event occurring in a particular geographic region, wherein the
3 identifying is based on disseminated information containing public commentary in the
4 particular geographic region;
5 identifying attributes that are related to the event;
6 extracting sentiment words relating to the identified event from the
7 disseminated information, wherein the extracted sentiment words are in a local
8 language of the particular geographic region; and
9 generating a sentiment trend visualization that depicts a trend of determined
10 sentiments of at least a particular one of the identified attributes, wherein the
11 determined sentiments are based on the sentiment words for at least the particular
12 attribute.
- 1 2. The method of claim 1, wherein identifying the attributes comprises selecting
2 the attributes from a collection of a larger number of candidate attributes according
3 to selection criteria.
- 1 3. The method of claim 1, wherein the sentiment trend visualization shows
2 sentiments of the at least one attribute over time.

1 4. The method of claim 3, further comprising:

2 presenting for display a list of the identified attributes;

3 receiving selection of a second one of the identified attributes, where the

4 second attribute is different from the particular attribute; and

5 changing the sentiment trend visualization to show sentiments of the second

6 attribute over time.

1 5. The method of claim 1, wherein the disseminated information includes data
2 records, the method further comprising:

3 receiving sentiment scores associated with the particular attribute, where the

4 sentiment scores express sentiments associated with the particular attribute in

5 corresponding ones of the data records, and where the sentiment scores comprise

6 sentiment scores at different times;

7 aggregating corresponding sentiment scores within each time interval of a

8 sequence of time intervals, where each of the time intervals is associated with a

9 respective aggregate score computed by the aggregating of corresponding

10 sentiment scores,

11 wherein the sentiment trend visualization of the particular attribute comprises

12 a graph of the aggregate scores.

- 1 6. The method of claim 1, further comprising:
2 generating a second visualization that includes plural bars for respective
3 plural ones of the identified attributes,
4 wherein each of the plural bars includes an arrangement of cells assigned
5 respective visual indicators representing corresponding different sentiments.
- 1 7. The method of claim 6, wherein the arrangement of cells in each of the bars is
2 based on sentiment scores associated with data records represented by the cells,
3 wherein the data records contain the disseminated information.
- 1 8. The method of claim 7, wherein each of the bars includes a first portion
2 containing cells associated with positive sentiment scores, and a second portion
3 containing cells associated with negative sentiment scores.
- 1 9. The method of claim 8, wherein each of the bars includes a third portion
2 containing cells associated with neutral sentiments scores.
- 1 10. The method of claim 6, wherein the visual indicators include different colors
2 corresponding to respective different sentiment scores.
- 1 11. An article comprising at least one machine-readable storage medium storing
2 instructions that upon execution cause a system having a processor to perform a
3 method according to any of claims 1-10.

1 12. A system comprising:

2 storage media to store data records containing disseminated information
3 including public commentary about an event in a local geographic region; and
4 at least one processor to:

5 identify the event in the disseminated information;

6 select a set of attributes from among candidate attributes in the
7 disseminated information, wherein the set of attributes is related to the event;

8 extract sentiment words relating to the identified event from the
9 disseminated information, where the extracted sentiment words are in a local
10 language of the local geographic region, and where the extracted sentiment words
11 express sentiments with respect to the attributes in the set; and

12 generate a sentiment trend visualization that depicts a trend of
13 determined sentiments of plural ones of the attributes in the set, wherein the
14 determined sentiments are based on the sentiment words for the plural attributes.

1 13. The system of claim 12, wherein the sentiment trend visualization includes a
2 plurality of bars corresponding to the respective plural attributes, where each of the
3 plurality of bars includes an arrangement of cells representing respective ones of the
4 data records, and where each of the cells has a visual indicator assigned based on a
5 sentiment score assigned to the respective attribute.

1 14. The system of claim 12, wherein the identified event is a previously unknown
2 event.

- 1 15. The system of claim 12, wherein the at least one processor is to further:
2 generate a second visualization that depicts a trend over time of a particular
3 one of the attributes in the set, wherein the second visualization includes a first
4 graph representing positive sentiments, and a second graph representing negative
5 sentiments.

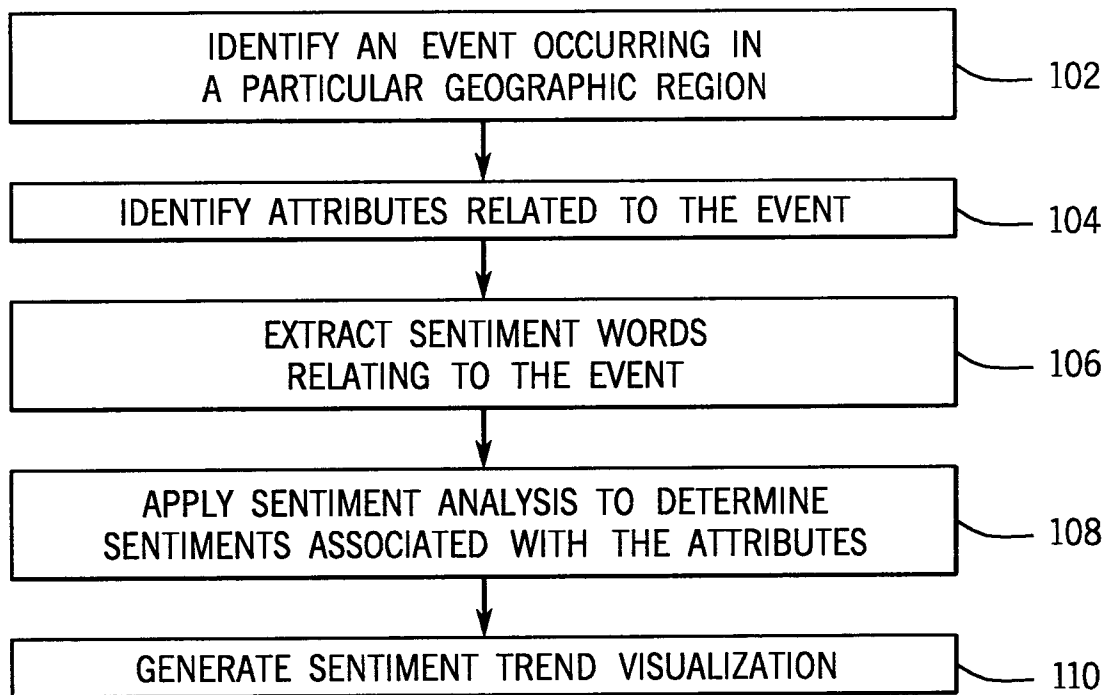


FIG. 1

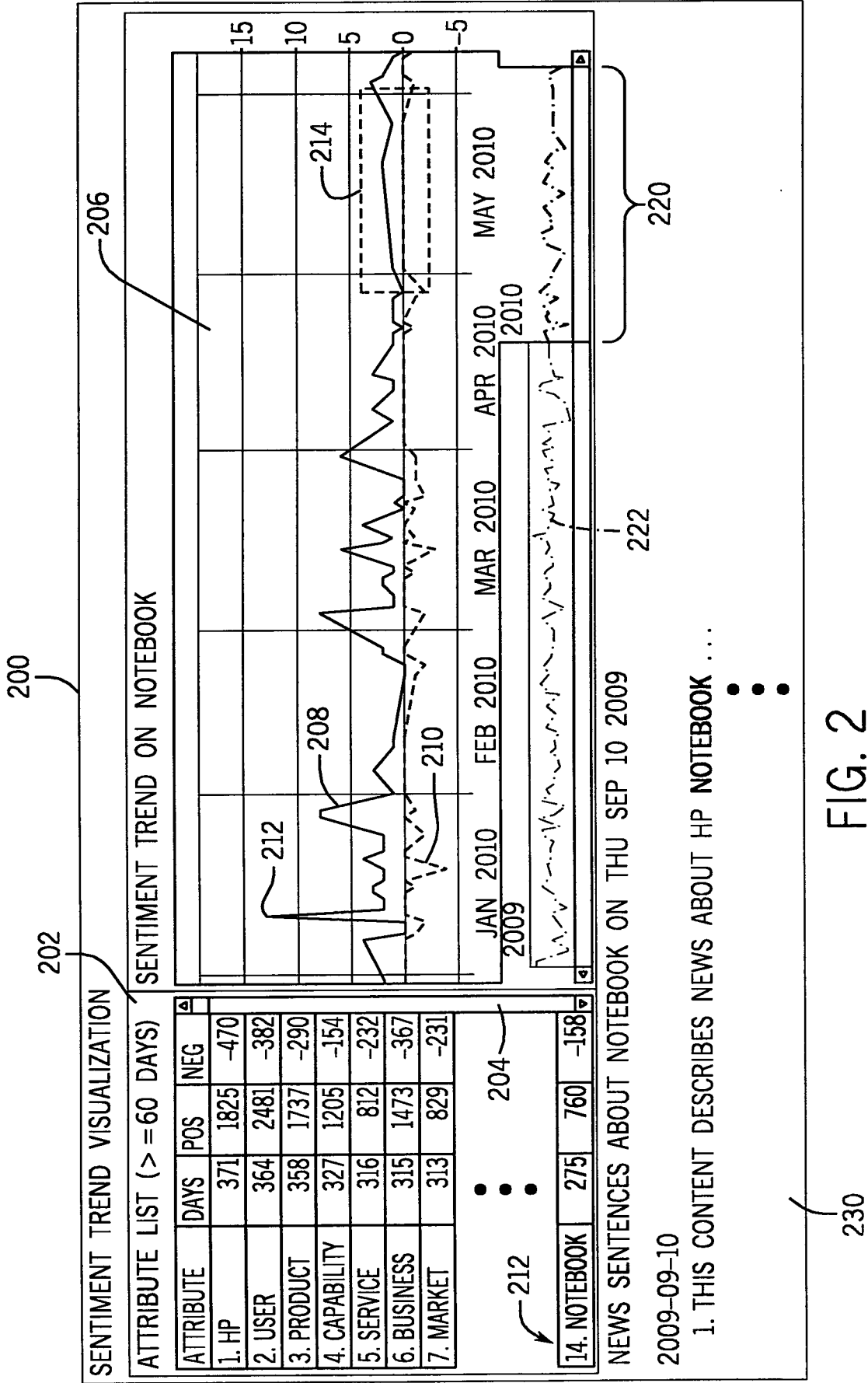


FIG. 2

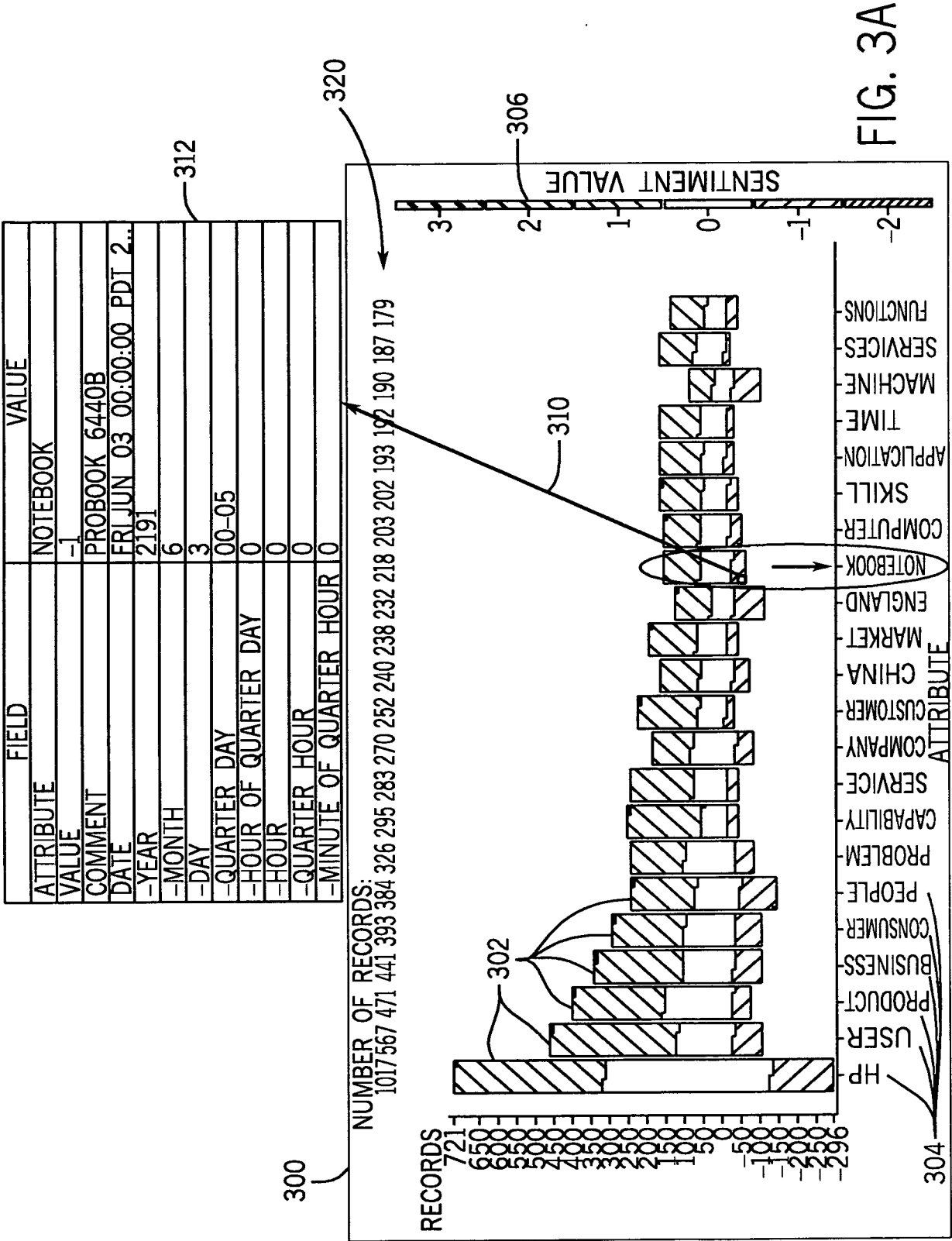
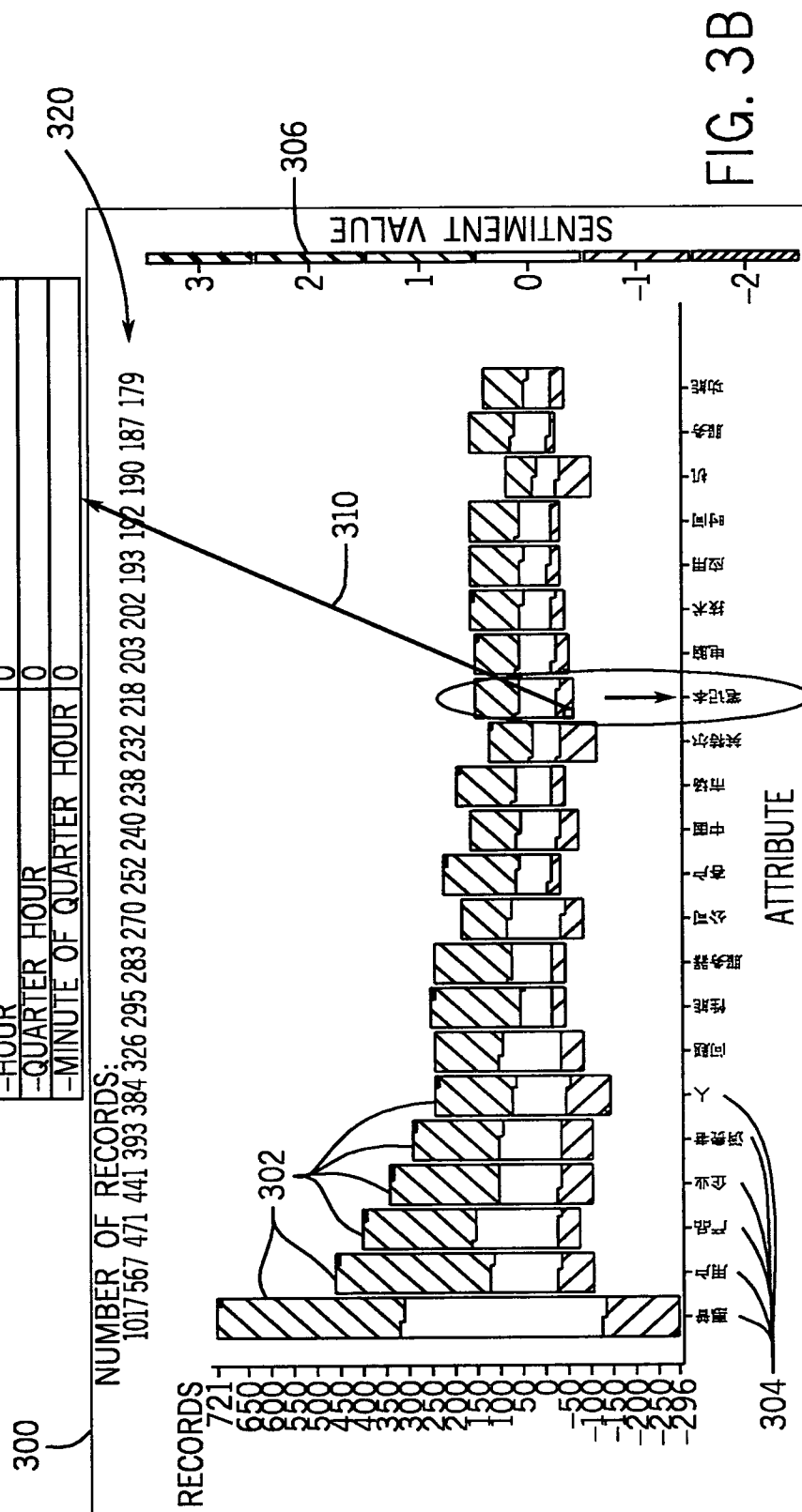


FIG. 3A

FIELD	VALUE
ATTRIBUTE	笔记本
VALUE	-1
COMMENT	惠普 PROBOOK 6440B (W...
DATE	FRI JUN 03 00:00:00 PDT 2..
-YEAR	2191
-MONTH	6
-DAY	3
-QUARTER DAY	00-05
-HOUR OF QUARTER DAY	0
-HOUR	0
-QUARTER HOUR	0
-MINUTE OF QUARTER HOUR	0



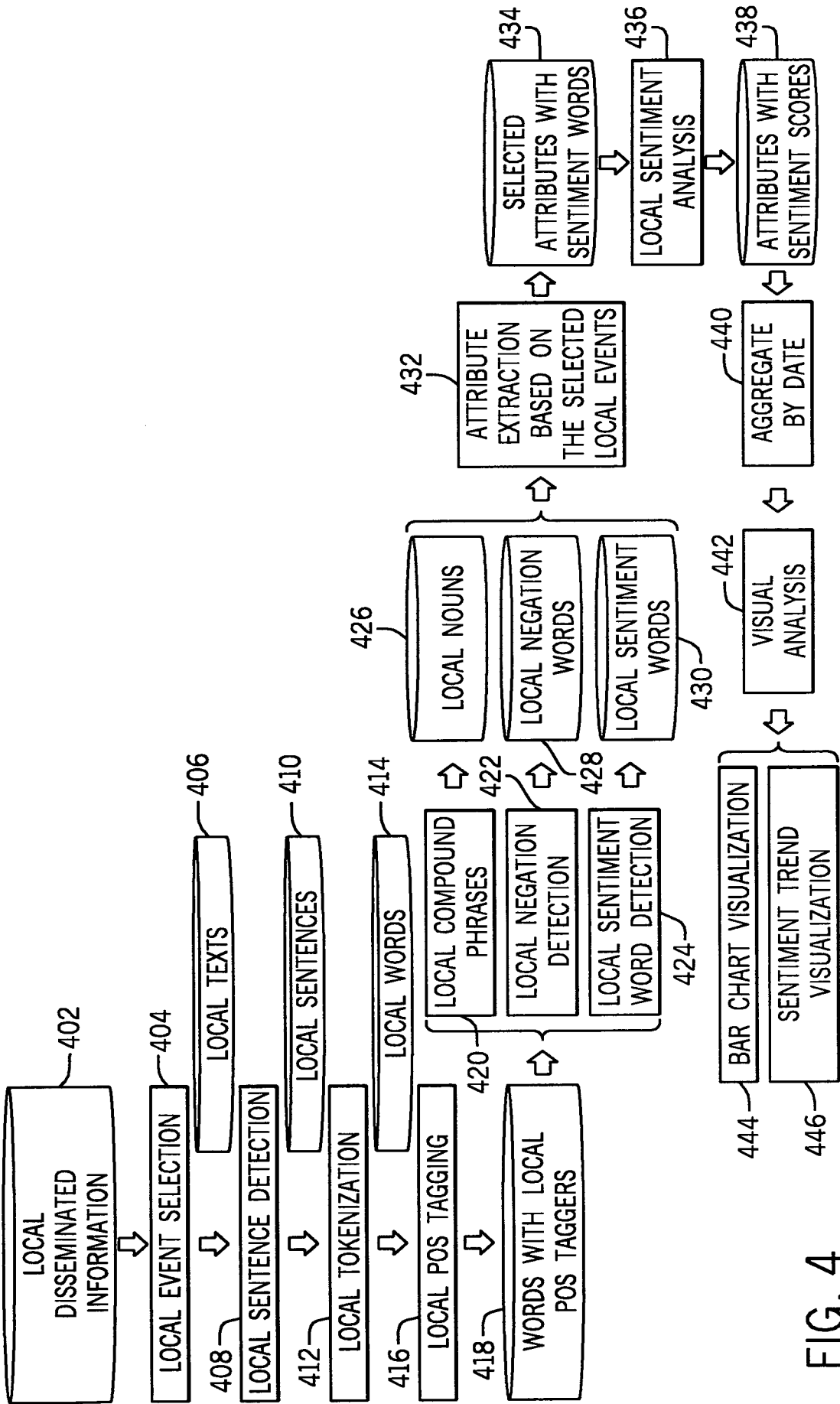


FIG. 4

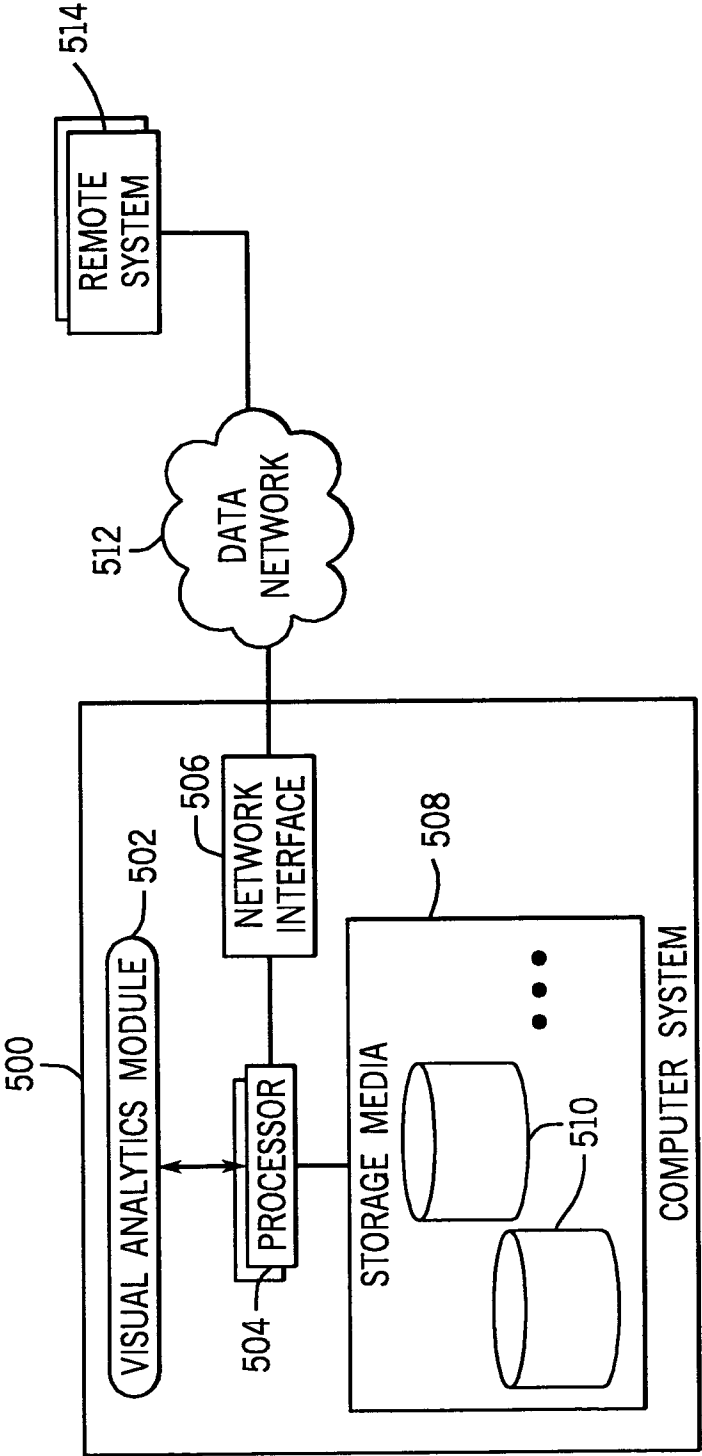


FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2011/000957

A. CLASSIFICATION OF SUBJECT MATTER

G06F17/30(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)

IPC: G06F

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

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

DWPI, EPODOC, USTXT, IEEE, GOOGLE: sentiment?, visualiz+, geograph???, attribute?, value?, local w language, sentiment, trend

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2010/0119053 A1 (BUZZIENT INC.) 13 May 2010 (13.05.2010) claims 1, 52; paragraphs 0003-0004, 0049-0060, 0065-0113 in the description; figures 1-35	1-15
A	US 2008/0215571 A1 (MICROSOFT CORPORATION) 04 Sep. 2008 (04.09.2008) the whole document	1-15
A	US 7600017 B2 (BUZZMETRICS LTD.) 06 Oct. 2009 (06.10.2009) the whole document	1-15
A	US 2008/0162238 A1 (RAJESH Venkat Subbu et al.) 03.Jul. 2008 (03.07.2008) the whole document	1-15
A	US 2008/0033587 A1 (KEIKO Kurita et al.) 07 Feb. 2008 (07.02.2008) the whole document	1-15

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

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search
02 December 2011 (02.12.2011)Date of mailing of the international search report
09 Feb. 2012 (09.02.2012)Name and mailing address of the ISA/CN
The State Intellectual Property Office, the P.R.China
6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China
100088
Facsimile No. 86-10-62019451

Authorized officer

LI, Jipei

Telephone No. (86-10)62414399

International application No.
PCT/CN2011/000957

Form PCT/ISA /210 (patent family annex) (July 2009)