

determined rules. The opinion rating subsystem further analyzes the message information and assesses an opinion rating according to a plurality of pre-determined linguistic and associative rules. The central data store of the present invention comprises one or more non-volatile memory devices for storing electronic data including, for example, message information, results of analyses performed by the system and a plurality of other information used in the present invention. In a preferred embodiment, the central data store further comprises a relational database system for storing the information in the non-volatile memory devices. The data analysis system of the present invention may comprise an objective data collection subsystem, an analysis subsystem, and a report generation subsystem. The objective data collection subsystem interfaces with a plurality of pre-determined objective data sources to collect data which may be used to establish trends and correlation between real-world events and the communication expressed in the various electronic discussion forums. The analysis subsystem performs the analysis of the objective data and message information described above. The report generation subsystem generates reports of the analysis to end-users. The reports may comprise pre-determined query results presented in pre-defined report formats or, alternatively may comprise ad hoc reports based on queries input by an end-user.

Brief Summary Text (12):

The method of the present invention comprises one or more of the steps of collecting a plurality of message information from a plurality of pre-determined electronic discussion forums; storing the plurality of message information in a central data store; categorizing the message information according to a plurality of pre-determined rules; assigning an opinion rating to the plurality of message information based on a plurality of pre-determined linguistic patterns and associative rules; collecting a plurality of objective data from a plurality of objective data sources; analyzing the message information and the objective data to identify trends in the pattern of behavior in pre-determined markets and the roles of participants in electronic discussion forums; and generating reports for end-users based on the results of the analyses performed by the present invention.

Description Paragraph (15):

Message Body--the portion of an electronic message comprising the pseudonym's contribution to the electronic discussion. The Message Body generally comprises the data, opinions or other information conveyed in the electronic message, including attached documents or files. Header Information--the portion of an electronic message not including the message body. Header Information generally comprises the transmission path and time/date stamp information, the message sender's information, the message identification number ("message ID"), the subject. Buzz Level--for a community, a measure of activity within the community, as determined by the number of distinct pseudonyms posting one or more messages over a given time frame. Connectivity--for a community, a measure of its relatedness with other communities, as determined by the number of other communities in which a community's participants concurrently participate. Actor--descriptive name of the role that a pseudonym plays in the social networks of communities. Actors can be further classified according to the following definitions: Initiator--a pseudonym that commences a discussion, i.e., one that posts the first message leading to subsequent responses forming a dialog on a particular subject. Moderator--a pseudonym that ends a discussion, i.e., one that posts the final message closing the dialog on a particular subject. Buzz Accelerator--a pseudonym whose postings tend to precede a rising buzz level in a community. Buzz Decelerator--a pseudonym whose postings tend to precede a falling buzz level in a community. Provoker--a pseudonym that tends to start longer discussion threads; different from buzz accelerators in that the metric is one discussion thread, not the community's overall discussion level. Buy Signaler--a pseudonym whose postings on a topic tend to precede a rising market for that topic. Sell Signaler--a pseudonym whose postings on a topic tend to precede a falling market for that topic. Manipulator--a pseudonym with little posting history except as Manipulators, whose combined

postings on one topic elevate the buzz level in the absence of external confirming events. Connector--a pseudonym who posts on a high number of topics or a high number of communities. Market Mood--a positive/negative market forecast derived from analysis of the patterns of actors' behavior.

Description Paragraph (17):

In a preferred embodiment, the present invention is implemented using a system architecture as shown in FIG. 1. The system architecture comprises electronic discussion data system 10, central data store 20, and analysis system 30. Electronic discussion data system 10 interfaces via network 4 with selected electronic discussion forums 6 to collect electronic messages and analyze intrinsic data comprising the messages according to one aspect of the present invention. Network 4 may be any communications network, e.g., the Internet or a private intranet, and may use any suitable protocol for the exchange of electronic data, e.g., TCP/IP, NNTP, HTTP, etc. Central data store 20 is a repository for electronic messages collected, objective data gathered from external sources and the results of the various analyses or reports produced by the system and method of the present invention. Central data store 20 may be implemented using any suitable relational database application program, such as, e.g., Oracle, Sybase and the like. Data analysis system 30 receives input from selected objective data sources for use in analyzing and quantifying the importance of the electronic discussion messages collected, and provides computer programming routines allowing end-users 9 to generate a variety of predefined and ad hoc reports and graphical analyses related to the electronic discussion messages. Each of the main systems comprising the system architecture of the present invention is described in more detail below.

Description Paragraph (18):

Central Data Store

Description Paragraph (19):

Central data store 20 comprises one or more database files stored on one or more computer systems. In a preferred embodiment, central data store 20 comprises message information database 22, topics database 23, objective data database 24, forum configuration database 25, analysis database 26 and reports database 27, as shown in FIG. 1. Message information database 22 comprises the message information collected by message collection subsystem 12. In a preferred embodiment, message information database 22 comprises: a message ID, i.e., a number or other string that uniquely identifies each message; sender information, i.e., the pseudonym, e-mail address or name of each message's author; a posting time and date for each message (localized to a common time zone); a collection time and date for each message; a subject field, i.e., the name of the thread or subject of each message; the message body for each message; an in-reply-to field, i.e., the message ID of the message to which each message was a reply; and the source of the message.

Description Paragraph (20):

The function and content of central data store 20's database files 23 27 are described in subsequent sections below.

Description Paragraph (21):

Electronic Discussion Data System

Description Paragraph (22):

As discussed above, electronic discussion data system 10 gathers certain messages and analyzes them according to the intrinsic information comprising the messages. Electronic discussion data system 10 comprises three subsystems: message collection subsystem 12, message categorization subsystem 14 and opinion rating subsystem 16. Message collection subsystem 12 collects message information from data sources and stores the information in central data store 20 for later analysis. Message categorization subsystem 14 extracts information about each message in central data store 20 and categorizes the messages according to a plurality of pre-defined

topics. The subsystem analyzes all aspects of each message and determines if the message is relevant to one or more of the topics that the system is currently tracking. A relevancy ranking for each message is stored in central data store 20 for each topic indicating the strength of the message's relation to each topic. Further analysis of the collected message information is carried out by opinion rating subsystem 16 to determine whether the message conveys a positive, neutral or negative opinion regarding the related topic. Each of the subsystems of electronic discussion data system 10 are described in more detail below.

Description Paragraph (24):

Message collection subsystem 12 collects electronic message information from the designated electronic discussion forums and passes the collected messages to central data store 20 and to message categorization subsystem 14, as shown in FIG. 1. The collected messages comprise records stored in message information database 22 in central data store 20. Database 22 comprises records including message header information and the message body. In a preferred embodiment, each field comprising message header information comprises a separate field of a record in database 22. The architecture used in a preferred embodiment of the present invention for implementing message collection subsystem 12 is shown in the schematic diagram in FIG. 2. This architecture supports multiple configurations for data collection and is highly scalable for gathering large or small amounts of message information. FIG. 2 illustrates some of the configurations that may be used in a preferred embodiment of message collection subsystem 12.

Description Paragraph (25):

As shown in FIG. 2, the message collection subsystem consists of several components that function together to collect information from electronic discussion forums 61 and 62 or discussion data files 63 and 64 on distributed networks 41 44. Although shown as separate discussion forums, data files and networks, it would be apparent to one skilled in the art that discussion forums 61 and 63 and data files 63 and 64 could be the same discussion forum or data file, and networks 41 44 could comprise a single distributed network, such as the Internet. Components of message collection subsystem 12 include message collector programs and message processor programs running on one or more computer systems. The computer systems used by message collection subsystem 12 comprise any suitable computers having sufficient processing capabilities, volatile and non-volatile memory, and support for multiple communications protocols. In a preferred embodiment, the computer systems used by message collection subsystem 12 comprise UNIX-based servers such as available from Sun Microsystems, or Hewlett-Packard and the like. All of the subsystem components can be replicated within a single computer system or across multiple computer systems for overall system scalability.

Description Paragraph (26):

In a preferred embodiment, message processor programs, e.g., message processor 121a and 121b, are in communication with database 22, which is part of central data store 20 (not shown in FIG. 2). In FIG. 2, the message processors and central data store are protected from unauthorized access by firewall security system 122. Other components of message collection subsystem 10 are located at various points in the architecture, as described below. As would be apparent to one of ordinary skill in the art, firewall 122 is provided for security and is not technologically required for operation of the present invention. Message processors 121a and 121b receive information from the message collectors and store the information in the database 22 for later processing. As shown in FIG. 2, message processors 121a and 121b may service more than one message collector program to facilitate processing of a large volume of incoming messages. Inbound messages are held in a queue on the message processors, allowing message processors 121a and 121b to receive many more messages from the message collectors than they can actually process for storing in database 22. This architecture allows the rapid collection of millions of messages from tens of thousands of discussion forums without excessive overloading of the computer systems.

Description Paragraph (27):

As is known in the art, each discussion forum or data file may have a unique message format. For example, an electronic message from one discussion forum may place the date field first, the message ID second, and the other header and body data last. A different discussion forum may choose to display the message ID first, followed by the pseudonym of the participant, and the message body. Moreover, each type of discussion forum has its own communications protocol. For example, the communications protocol for an interactive discussion forum (e.g., a chat session) is not the same as the communications protocol for USENET news groups. The message format and protocols need not be static, i.e., as discussion forums evolve, different data structures and protocols may be implemented. To accommodate such changes, each message collector receives configuration information from forum configuration database 25 in central data store 20, either directly or via the message processor systems. The configuration information indicates the data source, i.e., the discussion forum or discussion file, from which messages will be collected. The configuration information further comprises programming instructions tailored for each individual data source to allow the message collector program to communicate with the data source and extract and parse the message information. Accordingly, message collectors can support a wide variety of protocols utilized by discussion forums including, e.g., HTTP, NNTP, IRC, SMTP and direct file access. In a preferred embodiment, the general programming instructions are written the Java programming language with parsing instructions written in Jpython scripting language. By storing the configuration information in a centralized location, i.e., central data store 20, management of the message collectors is simplified. Accordingly, when the data structure for a particular discussion forum changes, the configuration information needs to be modified only once.

Description Paragraph (29):

As noted above, there are several ways to implement the architecture supporting message collection subsystem 12. In one implementation, message collector programs, shown in FIG. 2 as local message collectors 123a and 123b, are part of local area network ("LAN") 124 and are authorized access through firewall 122. Local message collector 123a interfaces through network 41 to collect messages from discussion forum 61 and local message collector 123b has direct access to discussion data file 63. The latter configuration may be implemented, e.g., if the operator of message collection subsystem 12 also hosts a community for message discussion forums. As shown in FIG. 2, a message collector may collect messages from multiple discussion forums. For example, as shown in FIG. 2, local message collector 123b also interfaces through network 41 to collect messages from discussion forum 61.

Description Paragraph (30):

In an alternative implementation, message collector programs, such as remote message collectors 125a and 125b, are run on external networks. As shown in FIG. 2, the remote message collectors are not part of LAN 124 and do not have direct access to the message processor programs running behind firewall 122. For security reasons, proxy servers 126a and 126b are used to interface with message processor 121b through firewall 122. Functionally, remote message collectors operate in the same manner as the local message collectors. That is, remote message collectors 125a and 125b receive configuration information from central data store 20 (via proxy servers 126a and 126b, respectively). Moreover, remote message collectors may collect messages from discussion forums over a network or directly from discussion data files, as shown in FIG. 2. Use of remote message collectors allows for geographic distribution and redundancy in the overall message collection subsystem architecture.

Description Paragraph (32):

Message categorization subsystem 14 analyzes the data collected from discussion forums and categorizes the messages into meaningful groupings, i.e., parent topics and topics, according to predefined rules as described below. In a preferred

embodiment, message categorization subsystem 14 retrieves message information from database 22 and topic information from central data store 20 and stores results of the categorization process in database 22. Alternatively, message categorization subsystem 14 may receive input directly from message collection subsystem 12 for immediate processing into categories.

Description Paragraph (33):

Topics database 23 comprises representations of real world topics that are being tracked and analyzed by the system and method of the present invention. FIG. 3 shows the hierarchical data structure used in a preferred embodiment of database 23. In a preferred embodiment, abstract root 231, shown in FIG. 3 as the top-level of the hierarchy, is not an actual topic stored in database 23 and is shown only to illustrate the hierarchy. Similarly, branches 232 234 are shown in FIG. 3 to conceptually show the relationship between topics stored in database 23. Accordingly, branch 232 indicates that some topics stored in database 23 may relate consumer entertainment, branch 233 indicates other topics relate to stock markets, and branch 234 may include other topics, such as, e.g., food, sports, technology adoption, and the like. As shown in FIG. 3, the hierarchy comprises one or more parent topics, such as parent topics 235 (related to books), parent topic 236 (related to movies), parent topic 237 (related to market indexes) and parent topic 238 (related to companies). Topics in the hierarchy are the last level, such as, topic 235a (Tears of the Moon), topic 235b (The Indwelling), topic 235c (Hot Six) and topic 235d (The Empty Chair). As shown in FIG. 2, topics 235a 235d are related to each other by parent topic 235 (books).

Description Paragraph (34):

In a preferred embodiment of the present invention, message categorization subsystem 14 assigns a relevance ranking for each topic to each message collected by message collection subsystem 12. The relevance ranking is determined based on a set of predefined rules stored in database 23 for each topic. The rules comprise a series of conditions defining information relevant to the topic, having an associated weighting to indicate the strength a particular condition should have in determining the overall relevance rank of the message with respect to the topic. Messages that need categorization are processed by message categorization subsystem 14 synchronously, i.e., the rules for each topic are applied to each message regardless of the relevance ranking for prior topics. The elements of each message, including subject, source, and content are processed against the conditions of each topic in the database. Based on the conditions that are satisfied and the weights of those conditions, a relevance rank for each topic is assigned to each message. As messages are processed, their relevance ranking for each topic is updated in message information database 22 in central data store 20.

Description Paragraph (39):

Opinion rating subsystem 16 extracts message information from database 22 in central data store 20 and assigns an opinion rating for each message by analyzing textual patterns in the message that may express an opinion. The textual patterns are based on linguistic analysis of the message information. For example, if the message body includes words such as "movie" and "awful" in the same sentence or phrase and the message had a high relevancy ranking for the topic "The Perfect Storm" the message may be expressing a negative opinion about the movie. Textual pattern analysis software, such as available from Verity Inc, of Mountain View, Calif., may be used to assign the opinion rating for each message. Such passive opinion polling is useful for market analysis without the need for individually interviewing active participants in a survey. Once the rating process is complete, the rating for each opinion processed is stored in database 22 in central data store 20.

Description Paragraph (40):

Data Analysis System

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.