US 20040006473A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2004/0006473 A1

Mills et al. (43) Pub. Date: Jan. 8, 2004

(54) METHOD AND SYSTEM FOR AUTOMATED CATEGORIZATION OF STATEMENTS

(75) Inventors: **Scott H. Mills**, Austin, TX (US); **Kurt M. Joseph**, Austin, TX (US); **John M. Martin**, Austin, TX (US); **Benjamin A. Knott**, Round Rock, TX (US); **Robert R. Bushey**, Cedar Park, TX (US); **Theodore B. Pasquale**, Austin, TX (US)

Correspondence Address:
BAKER BOTTS L.L.P.
PATENT DEPARTMENT
98 SAN JACINTO BLVD., SUITE 1500
AUSTIN, TX 78701-4039 (US)

(73) Assignee: **SBC Technology Resources, Inc.**, Austin, TX

(21) Appl. No.: **10/188,152**

(22) Filed: **Jul. 2, 2002**

Publication Classification

(51) **Int. Cl.**[7] **G10L 21/00**

(52) **U.S. Cl.** **704/270**

(57) **ABSTRACT**

A method and system for automating categorization of statements includes a categorization system having a plurality of rules to categorize the statements, a rule engine, and a category engine. The rule engine allows for the creation and storage of objective rules used to categorize the statements. The category engine automatically applies the rules to a list of statements in order to categorize the statements and automatically determines a category label for each statement. The category engine further creates an output file including each statement and the corresponding category label. The use of objective rules to categorize the statements allows for reliable and consistent categorization results and eliminates any subjectiveness in the categorization of the statements.

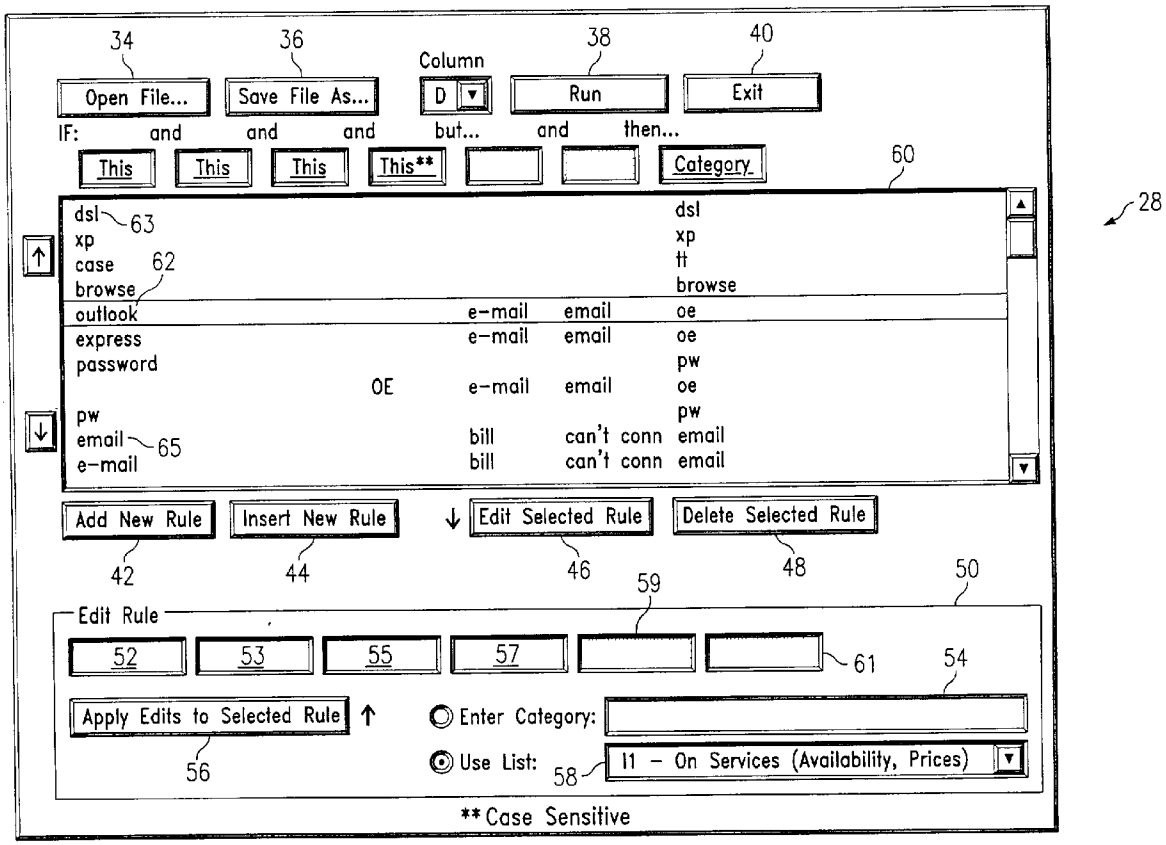


FIG. 1

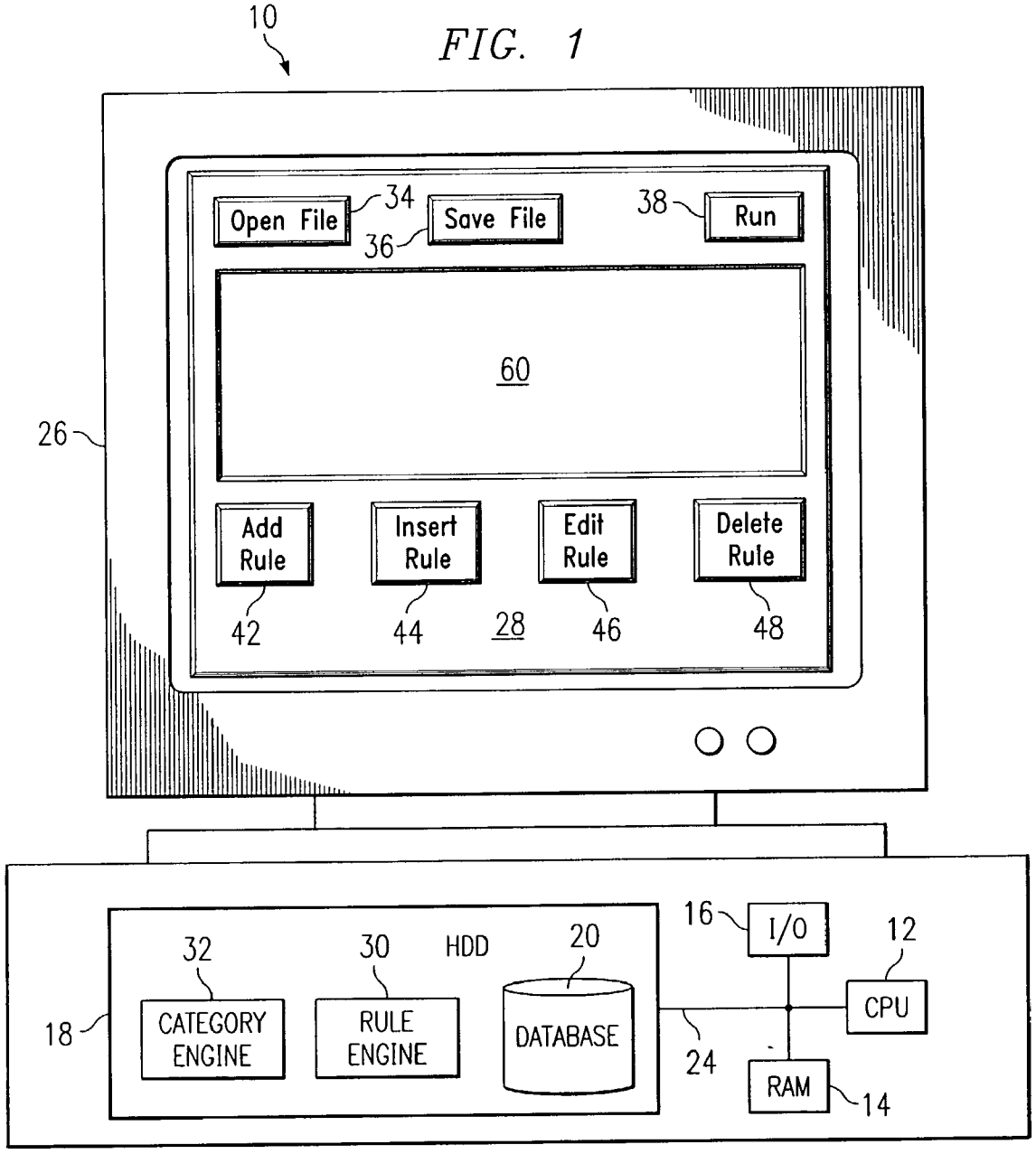
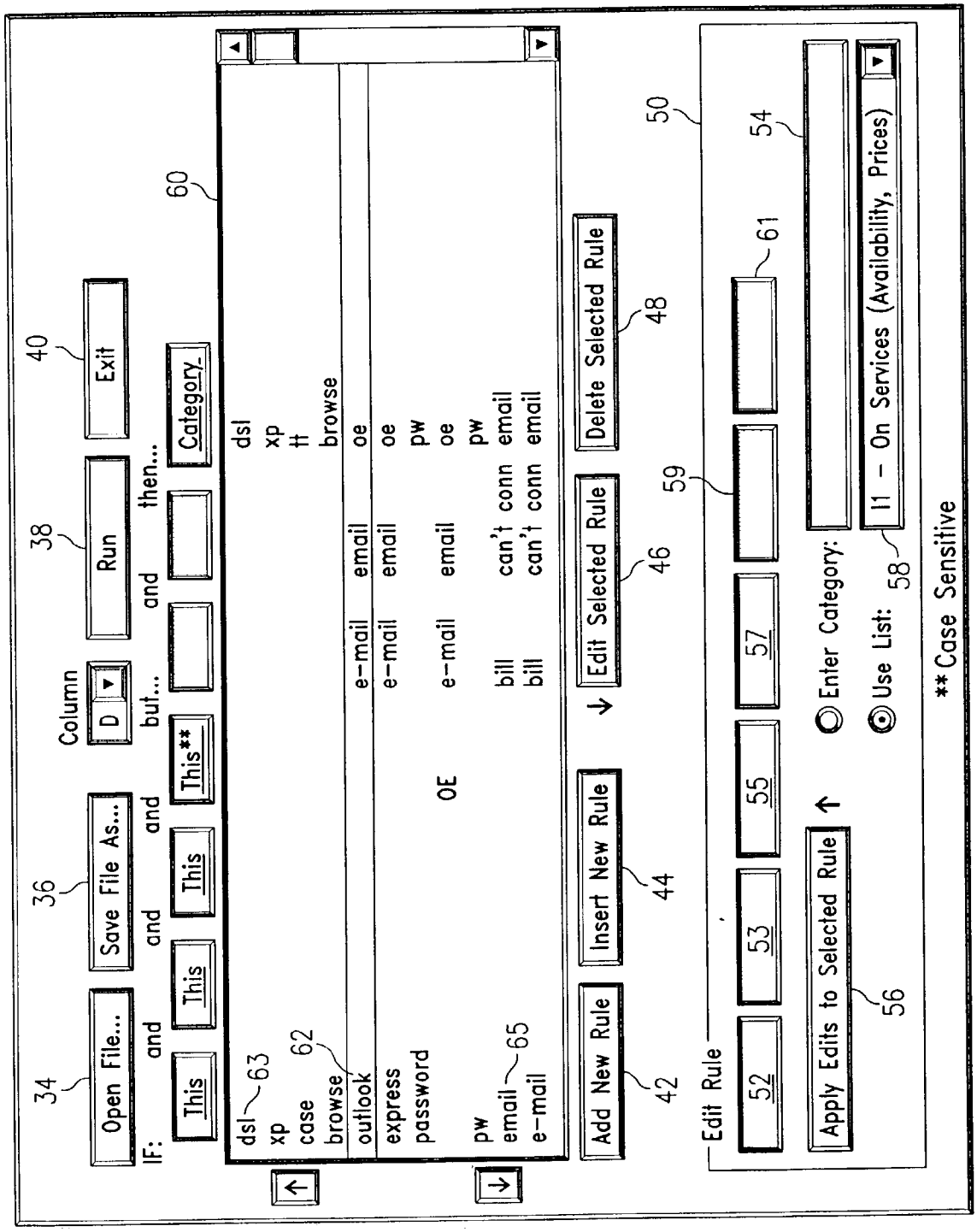

10
Open File
34
Save File
36
38
Run
60
26
Add Rule
Insert Rule
Edit Rule
Delete Rule
42
44
28
46
48
32
30
HDD
20
16
I/O
12
CPU
18
CATEGORY ENGINE
RULE ENGINE
DATABASE
24
RAM
14

FIG. 2


28
34
Open File...
36
Save File As...
40
Exit
38
Run
Column
D
Category
IF:
This
and
This
and
This
and
This**
but...
and
then...
dsl
xp
case
browse
outlook
express
password
pw
email
e-mail
62
63
65
60
dsl
xp
tt
browse
e-mail email oe
e-mail email oe
pw
OE e-mail email oe
pw
bill can't conn email
bill can't conn email
Add New Rule
42
Insert New Rule
44
Edit Selected Rule
46
Delete Selected Rule
48
50
Edit Rule
52
53
55
57
59
61
54
Apply Edits to Selected Rule
56
Enter Category:
Use List:
58
11 - On Services (Availability, Prices)
** Case Sensitive

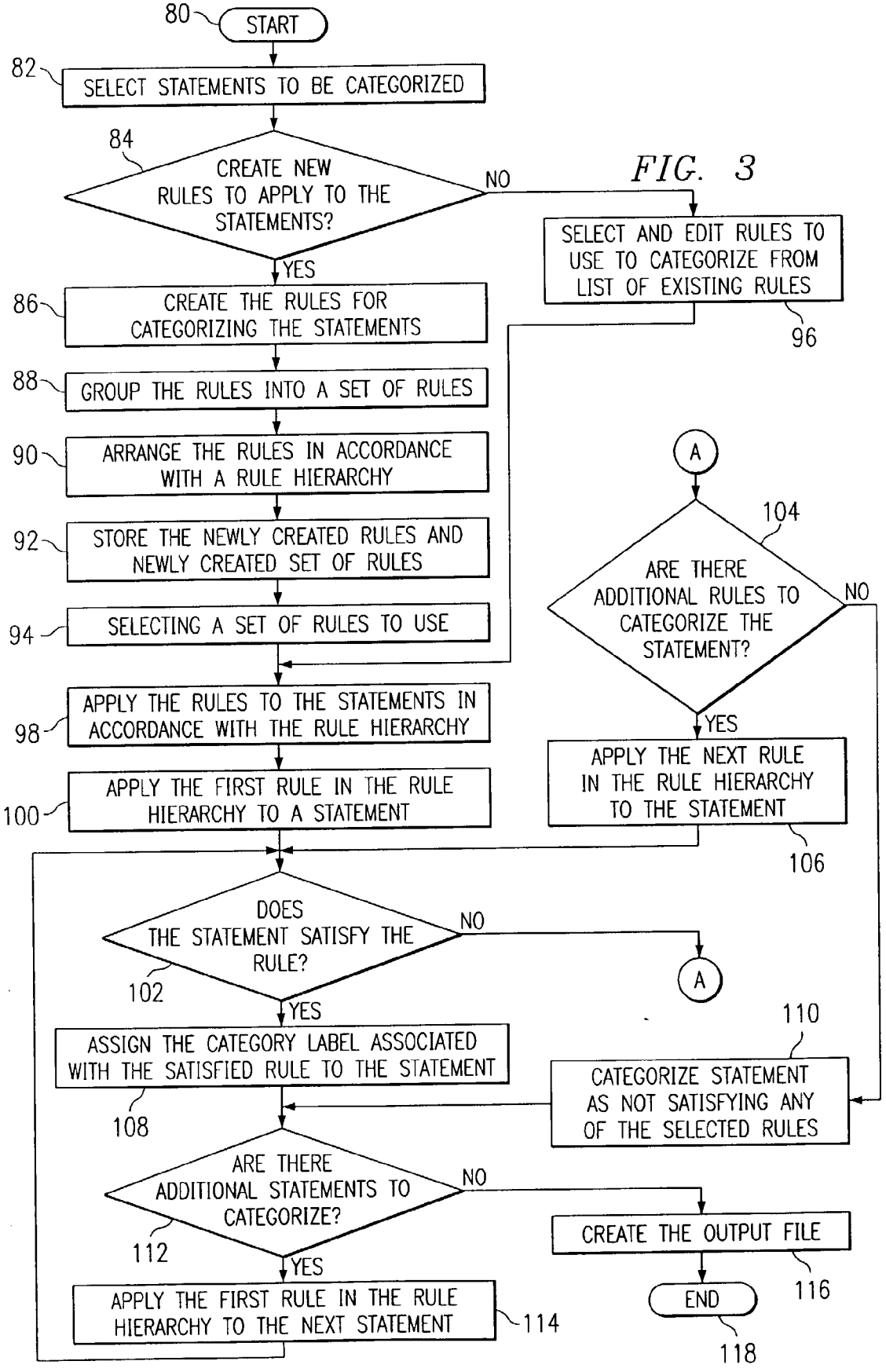

80
START
82
SELECT STATEMENTS TO BE CATEGORIZED
84
CREATE NEW RULES TO APPLY TO THE STATEMENTS?
NO
YES
SELECT AND EDIT RULES TO USE TO CATEGORIZE FROM LIST OF EXISTING RULES
96
86
CREATE THE RULES FOR CATEGORIZING THE STATEMENTS
88
GROUP THE RULES INTO A SET OF RULES
90
ARRANGE THE RULES IN ACCORDANCE WITH A RULE HIERARCHY
92
STORE THE NEWLY CREATED RULES AND NEWLY CREATED SET OF RULES
94
SELECTING A SET OF RULES TO USE
98
APPLY THE RULES TO THE STATEMENTS IN ACCORDANCE WITH THE RULE HIERARCHY
100
APPLY THE FIRST RULE IN THE RULE HIERARCHY TO A STATEMENT
A
104
ARE THERE ADDITIONAL RULES TO CATEGORIZE THE STATEMENT?
NO
YES
APPLY THE NEXT RULE IN THE RULE HIERARCHY TO THE STATEMENT
106
DOES THE STATEMENT SATISFY THE RULE?
NO
A
102
YES
110
ASSIGN THE CATEGORY LABEL ASSOCIATED WITH THE SATISFIED RULE TO THE STATEMENT
CATEGORIZE STATEMENT AS NOT SATISFYING ANY OF THE SELECTED RULES
108
ARE THERE ADDITIONAL STATEMENTS TO CATEGORIZE?
NO
YES
112
CREATE THE OUTPUT FILE
116
APPLY THE FIRST RULE IN THE RULE HIERARCHY TO THE NEXT STATEMENT
114
END
118

FIG. 3

METHOD AND SYSTEM FOR AUTOMATED CATEGORIZATION OF STATEMENTS

TECHNICAL FIELD OF THE INVENTION

[0001] The present invention relates generally to information processing and management, and more specifically relates to a method and system for automated categorization of statements.

BACKGROUND OF THE INVENTION

[0002] Customers often call a company call center or access a company's web page with problems or questions about a product or service or to alter the service or product. When calling, a customer often speaks to a customer service representative (CSR) or interacts with an interactive voice response (IVR) system and explains the purpose of the inquiry in the first statement made by the customer whether that be the first words spoken by the customer or the first line of text from a web site help page or an email. These statements made by customers are often referred to as opening statements and are helpful in quickly determining the purpose of a customer's inquiry.

[0003] Some companies track and classify the opening statements provided by customers in order to better provide customer interfaces that are in accordance with the way customers think. Companies typically manually track the statements provided by the customers and manually categorize the statements in order to determine frequencies of occurrence with respect to how often customers inquire about certain products and/or services. Manually categorizing the statements is a difficult task that is costly, time consuming, and subjective in that the categorizations may vary based on each person's personal opinion as to how a statement should be classified.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] A more complete understanding of the present embodiments and advantages thereof may be acquired by referring to the following description taken in conjunction with the accompanying drawing, in which like reference numbers indicate like features, and wherein:

[0005] **FIG. 1** depicts a block diagram of a system for automating the categorization of statements;

[0006] **FIG. 2** illustrates an example graphical user interface; and

[0007] **FIG. 3** depicts a flow diagram of a method for the automated categorization of statements.

DETAILED DESCRIPTION OF THE INVENTION

[0008] Preferred embodiments of the present invention are illustrated in the figures, like numerals being used to refer to like and corresponding parts of the various drawings.

[0009] Many companies that have customer service programs and/or call centers, such as telephone companies, Internet service providers, and credit card companies, often track statements made by customers when the customers contact the company with problems or questions about a product or service or to alter a product or service. When a customer calls a service number and speaks to a customer service representative (CSR), the customer typically tells the CSR the purpose of the call in the first substantive statement the customer makes. Alternatively, a customer may contact a company via the company web site or email and generally the first substantive statement made in the email or web site response includes the customer's purpose for contacting the company. These initial statements containing the purpose of the customer's call are often referred to as opening statements.

[0010] These opening statements can be used by companies to better design web sites, interactive voice response (IVR) systems, and any other customer interfaces between a company and the customers. One effective way to design an IVR system or a web site interface is to analyze the scripts of incoming calls or emails to a customer support center or call center to locate the opening statements and identify the purpose of each call or email by classifying or categorizing each opening statement. Once categorized, a frequency report can be created that details how often customers are calling with specific problems or questions about specific products or services. For example, a telephone company may want to know how many customers are calling or emailing about a problem with their bill or to add a new product to their telephone service. Once a company knows the frequency of customer complaints and questions, an IVR system can be designed that incorporates the frequencies so that customers calling with common problems, complaints, or questions can be serviced quickly and efficiently. For example, a company would be able to determine that of the 5,000 service calls received in one month, what percentage of the calls were about particular topics and also rank the reasons why the customers called or emailed the customer support.

[0011] In order to maximize the utilization of the statements given by the customers in a customer interface design, a company therefore needs to track and categorize the statements. Typically, companies have manually tracked and manually categorized opening statements. The company manually tracks each call and manually records and transcribes each opening statement spoken to a CSR or received via email and then creates a list of opening statements. An employee of the company then sits and reads the long list of opening statements with a list of categories in front of him/her and assigns a category label to each opening statement. This has proved to be a very time consuming and costly process because one or more people manually examining every opening statement and deciding how to categorize the statement in accordance with multiple category labels requires a large amount of employee time which is expensive and would be better utilized in a revenue generating task.

[0012] In addition to the cost and man-power required for the manual categorization of opening statements, there is also a subjective element to the manual categorization of opening statements which affects the reliability of the categorization results. The category labels used to manually categorize the opening statements are generally designed to be objective but when applied by a person, the person's subjective thinking and opinions affect how they categorize the opening statements. For instance, an opening statement such as "I am calling about my bill for the charges for Call Waiting" may be categorized by one person as a billing inquiry and another person as a call waiting inquiry. There-

fore, even though multiple people may use the same category labels to categorize the opening statements, they might categorize the same opening statement differently because the categorization is partly a matter of opinion. This human opinion factor and subjectiveness creates an inconsistency in the categorization data and frequency reports that results in unreliable data and a customer interface design that is not optimized with respect to the opening statements and the way customers think.

[0013] By contrast, the example embodiment described herein allows for the automated categorization of statements. Additionally, the example embodiment allows for the creation of objective rules to categorize the statements which results in reliable and consistent categorization data. Time and money is saved because people are no longer manually looking through lists of statements trying to categorize the statements using only category labels. Therefore, employees' time may be better utilized in revenue generating projects. Furthermore, the objective rules for categorizing the statements eliminate the subjective aspect of the categorization scheme allowing for the same statement to be categorized with the same category label as long as the same set of rules are used to categorize the statements. This results in consistent and reliable categorization and frequency data which can be used in the design and creation of customer interfaces that reflect the customers' view of how the interface should operate.

[0014] Referring now to **FIG. 1, a** block diagram depicts categorization system **10** for automating the categorization of statements. In the example embodiment, categorization system **10** may include respective software components and hardware components, such as processor **12**, memory **14**, input/output ports **16**, hard disk drive (HDD) **18** containing database **20**, and those components may work together via bus **24** to provide the desired functionality. The various hardware and software components may also be referred to as processing resources. Categorization system **10** may be a personal computer, a server, or any other appropriate computing device. Categorization system **10** may further include display **26** for presenting graphical user interface (GUI) **28** and input devices such as a mouse and a keyboard. Categorization system **10** also includes rule engine **30** and category engine **32**, which reside in memory such as hard disk drive **18** and are executable by processor **12** through bus **24**.

[0015] Categorization system **10** allows for the development of a one or more rules for the categorization of statements which are then applied to a list of statements in order to determine a category label for each statement. Display **26** presents GUI **28** which allows for the creation and editing of the rules and for the categorization of the statements. Shown in **FIG. 1** is an example GUI **28** with GUI **28** illustrated in greater detail in **FIG. 2**. GUI **28** includes a plurality of buttons that allow the user to access and control the operation of rule engine **30** and category engine **32** and also display the rules that are used to categorize the statements.

[0016] **FIG. 3** depicts a flow diagram of a method for the automated categorization of statements. The method begins at step **80** and at step **82** a user selects the statements to be categorized. Before categorization system **10** can automatically categorize the statements, the user must have one or more statements to categorize and load the list of statements into categorization system **10**. The statements may be opening statements as defined above, written statements from a training session, survey responses, search statements from a web site or pop-up window, statements evaluating a customer's experience and satisfaction in a test environment, or any other appropriate response to an open-ended question that can be analyzed using content text analysis.

[0017] Typically, the statements are recorded, transcribed, configured in a format that can be understood by categorization system **10**, and then placed in a text file which may be stored in database **20**. Because there may be more than one list of statements and therefore more than one text file, the user chooses what list of statements to categorize by selecting a text file using open file button **34**. Open file button **34** allows the user to view all the available files containing statements and then select the file containing the list of statements to be categorized. Once the list of statements has been selected, categorization system **10** reads the list of statements from database **20**.

[0018] After the selection of the statements to be categorized, at step **84** the user decides whether to use rule engine **30** to create new rules to categorize the statements or use existing rules already stored in database **20** to categorize the statements. If at step **84** the user decides to create new rules, then at step **86** the user accesses rule engine **30** to create new rules. New rules are desirable when there have been new products or services recently made available to the customers and the existing rules do not reflect these new products or services or when the statements are from a new domain not covered by the existing rules, such as survey responses where all the existing rules pertain to statements from customer service call centers.

[0019] The user utilizes rule engine **30** and rule creation screen **50** to create new rules and then edit the newly created rules. Creation of the rules involves the use of four include boxes **52**, **53**, **55**, and **57** and two exclude boxes **59** and **61**. In alternate embodiments, there may be more or less than four include boxes and more or less than two exclude boxes. The user inputs combinations of words and text strings that should be included in the statement in order for the statement to satisfy the rule include boxes **52**, **53**, **55**, and **57** and combinations of words and text strings that should not be in the statement in order for the statement to satisfy the rule in exclude boxes **59** and **61**. Each rule is also associated with a particular category label which the user enters in category label box **54**.

[0020] For example, a user may want to create a new rule to categorize statements with respect to the late payment of customer bills. Therefore "late" may be entered in include box **52**, "bill" may be entered in include box **53**, "paid" may be entered in exclude box **59**, and "labill" may be entered in category label box **54**. This allows for a rule that finds statements that contain the words "late" and "bill" but do not contain the word "paid." If a statement contains the words "late" and "bill" and does not include the word "paid," then the statement would be categorized with the category label "labill," meaning the purpose of the statement is to inquire about a late bill that has not yet been paid.

[0021] Once a user enters in the desired words or text strings in include boxes **52**, **53**, **55**, and **57** and exclude boxes **59** and **61**, the user selects apply rule button **56** and the rule appears in rule screen **60** and is available to be edited and

used to categorize the statements. The user may then repeat the above process to create as many rules as needed. In addition, alternate embodiments allow for rules where a noun in the singular form in include box **52** includes all forms of the noun (singular and plural) and a verb in the present tense in include box **52** includes all tenses and forms of that verb. This allows for a bigger hit rate when applying the rules to the statements since one rule is satisfied by a statements containing any form of the noun or verb and saves time because multiple rules are not required for each form of the noun or verb.

[**0022**] After the creation of the rules, at step **88** the user groups the rules into sets of rules. There may be different sets of rules for different applications or divisions of a company. For example, the marketing division may have a set of rules to categorize a list of statements while the product development division may have a different set of rules to categorize the same list of statements. This is because different users may be interested in different terms with respect to a list of statements. In addition, different sets of rules may also be necessary for different kinds of statements or statements from different domains. A user may use one set of rules to categorize opening statements from a call center and a different set of rules to categorize survey responses from a web survey questionnaire. Therefore, rule engine **30** allows for the rules to be grouped into different sets of rules with the name for each set of rules displayed in set box **58** and the sets of rules saved in database **20**. In addition, the user may group only newly created rules together in a group or group together newly created rules with existing rules when creating sets of rules.

[**0023**] At step **90**, the rules must be arranged in a rule order in accordance with a rule hierarchy enabling category engine **32** to apply the rules in the correct order thereby preventing inconsistent results. Typically the rule hierarchy is from specific rules to general rules but can be any other appropriate way of ordering the rules. For a specific to general rule hierarchy, category engine **32** applies the most specific rules first to a statement and then applies the more general rules if the statement does not satisfy any of the specific rules.

[**0024**] For example, a user wants to find both "phone" and "telephone" separately. A rule specifying "telephone" needs to be above the rule specifying "phone" in the rule hierarchy so that the "telephone" rule is applied to a statement before the "phone" rule is applied to a statement. If the "phone" rule is applied before the "telephone" rule, then when category engine **32** comes across a statement containing the word "telephone," category engine **32** will find "phone" in "telephone" and categorize the statement with the "phone" category label instead of the "telephone" category label and the statement will be incorrectly categorized. But if the "telephone" rule is placed above the "phone" rule in the rule hierarchy, then category engine **32** will find "telephone" in the statement, categorize that statement with the "telephone" category label and move on to the next statement without ever applying the "phone" rule. Therefore, the most specific rules need to be placed at the top of the rule hierarchy and the most general rules need to be placed at the very end or bottom of the rule hierarchy with a gradual gradient from specific to general in-between.

[**0025**] Once the rules have been grouped and ordered in a correct rule hierarchy, rule engine **30** stores the newly created rules, sets of rules, and rule hierarchy in database **20** at step **92** so that users and category engine **32** may later access the rules. After rule engine **30** saves the rules, at step **94** the user selects the rule or the set of rules that the user wants to have category engine **32** apply to the list of statements.

[**0026**] If at step **84** the user decides to not create any new rules but instead to use existing rules, then at step **96** the user selects and edits rules from the lists of existing rules stored in database **20**. Existing rules include rules that have already been created and saved by the process outlined above at steps **86** through **94**. If a user has already created a set of rules that has worked well in the past in categorizing statements, then the user may want to use these rules instead of creating new rules. The user selects from the list of rules in set box **58** and the rules from the selected set of rules appear in rule screen **60**. Once the rules appear in rule screen **60**, the user may edit an existing rule such as rule **62** by selecting it in rule screen **60** and clicking edit rule button **46**. The rule then appears in rule creation screen **50** and the user may modify include boxes **52**, **53**, **55**, and **57** and exclude boxes **59** and **61**. Once the user has a set of rules for category engine **32** to apply to the list of statements, the process continues to step **98**.

[**0027**] At step **98**, the user selects run button **38** and category engine **32** applies the selected rules to the list of statements in order to determine a category label for each statement. Category engine **32** cycles through the list of statements one statement at a time applying the rules to a statement until each statement satisfies a rule. Category engine **32** begins applying the rules to the list of statements at step **100** by applying the first rule in the rule hierarchy to the first statement in the list of statements. When category engine **32** applies the rules to the statements, category engine **32** strips the punctuation off the statements so that "bill," and "bill" do not appear as two different text strings.

[**0028**] At step **102**, category engine **32** determines if the statement satisfies the first rule. Category engine **32** determines if a statement satisfies a rule by searching the statement for the presence of particular text string combinations or words and the exclusion of other text string combinations or words. For instance, rule **63** is the highest rule in the rule hierarchy shown in rule screen **60**. Therefore, category engine **32** searches the first statement to see if the text string "dsl" is present in the first statement. If "dsl" is not present in the first statement, then the first statement does not satisfy rule **63**. If the statement does not satisfy the rule, then at step **104** category engine **32** checks to see if there are additional rules in the set of rules to apply to the statement. If there are additional rules to apply to the statement, then at step **106** category engine **32** applies the next rule in the rule hierarchy to the statement and the process returns to step **102** where category engine **32** determines if the statement satisfies this rule. Steps **102**, **104**, and **106** repeat until either the statement satisfies a rule at step **102** or until the statement does not satisfy any of the rules at step **102** and there are no more rules to apply to the statement at step **104**.

[**0029**] If the statement satisfies a rule at step **102**, then at step **108** category engine **32** assigns the category label associated with the satisfied rule to the statement. So if the statement contained the text string "dsl," then category engine **32** assigns the "dsl" category label to the statement.

But if the statement does not satisfy any of the rules at step **102** and there are no more rules left to apply at step **104**, then category engine **32** applies a catch-all rule to the statement and labels the statement with the catch-all category label at step **110**. The catch-all rule and category label is designed for statements that do not fit within any of the other rules. Category engine **32** labels the statement as catch-all so that the statement may be examined at a later date to determine if the statement really does not satisfy any of the rules or if there is a malfunction of categorizing system **10** which resulted in the statement not satisfying any of the rules. A high number of catch-all category labels may indicate that categorization system **10**, rule engine **30**, or category engine **32** are not operating correctly and require attention.

[**0030**] After category engine **32** assigns a category label to the statement at either step **108** or step **110**, at step **112** category engine **32** checks to see if there are additional statements in the list of statements that require categorization. If there are additional statements to be categorized at step **112**, then at step **114** category engine **32** selects the next statement to be categorized and applies the first rule in the rule hierarchy to the statement and then determines if the statement satisfies the rule at step **102**. Category engine **32** repeats steps **102-112** until category engine **32** determines at step **112** that there are no additional statements to be categorized.

[**0031**] Category engine **32** then cycles through the list of statements one statement at a time to determine a category label for each statement. When category engine **32** determines a category label for a statement, category engine **32** moves to the next statement. For instance, a statement to be categorized is "I cannot access my email account." Category engine **32** applies the first rule in rule screen **60**, rule **63**, to the statement. Category engine **32** applies rule **63** by searching the statement "I cannot access my email account" for the text string "dsl." Category engine **32** determines that the statement does not contain the text string "dsl" and therefore the statement does not satisfy rule **63**. Category engine **32** then applies each rule below rule **63** to the statement one rule at a time until the statement satisfies a rule. When category engine **32** gets to rule **65** and applies rule **65** to the statement, category engine **32** determines that the statement includes the text string "email" and does not include the text strings "bill" and "can't comm." Therefore, the statement satisfies rule **65** and category engine **32** assigns category label "email" to the statement and category engine **32** checks to see if there are any additional statements to categorize.

[**0032**] When there are no additional statements to be categorized, category engine **32** creates an output file at step **116** and the process ends at step **118**. The output file includes all the statements from the list of statements and each corresponding category label. An example output file with three statements is shown in Table 1. The output file allows a user to determine the frequency of occurrence for each category label and therefore determine which categories customers are calling the most about. Knowing which categories the customers are calling the most about allows for a customer interface design that takes into account the customers' way of thinking and is therefore easier to for the customer to use. The interface design that is easier for the customer to use allows the customer to accomplish their tasks in less time and a more efficient manner resulting in less company resources being used in servicing the customers and therefore lower costs for a company.

TABLE 1

Sample Output File

Statement	Category Label
I cannot access my email account	email
My DSL connection is slow	dsl
I have forgotten my password	password

[**0033**] Although the present invention has been described in detail, it should be understood that various changes, substitutions and alterations can be made hereto without the parting from the spirit and scope of the invention as defined by the appended claims.

What is claimed is:

1. A method for categorizing customer service opening statements, the method comprising:

collecting a plurality of opening statements to be categorized;

creating one or more rules for categorizing the opening statements;

grouping the rules into one or more sets of rules;

storing the sets of rules;

selecting one of the sets of rules to apply to the opening statements;

automatically applying the rules in accordance with a rule hierarchy to a list of the opening statements one opening statement at a time;

searching each opening statement for one or more text string combinations;

automatically determining a category label for each opening statement based upon the presence of one or more of the text string combinations;

assigning a category label to each opening statement when each opening statement first satisfies one of the rules; and

creating an output file including each opening statement and a corresponding category label.

2. A method for the automated categorization of statements, the method comprising:

creating one or more rules for categorizing the statements;

selecting one or more of the rules to apply to the statements;

automatically applying the rules to a list of the statements; and

automatically determining a category label for each statement based upon the rules.

3. The method of claim 2 wherein automatically applying the rules to a list of the statements comprises retrieving one or more of the rules to be applied to the statements.

4. The method of claim 2 wherein creating one or more rules comprises grouping the rules into one or more sets of rules.

5. The method of claim 2 further comprising creating an output file including each statement and a corresponding category label.

6. The method of claim 2 wherein automatically applying the rules to a list of the statements comprises applying the rules to the statements one statement at a time.

7. The method of claim 2 further comprising determining a rule hierarchy for applying the rules to the statements.

8. The method of claim 7 wherein automatically applying the rules to a list of the statements comprises applying the rules to the statements in a particular rule order in accordance with the rule hierarchy.

9. The method of claim 2 wherein automatically applying the rules to a list of the statements comprises searching each statement for one or more text string combinations.

10. The method of claim 2 wherein creating one or more rules comprises editing one or more existing rules.

11. The method of claim 2 wherein automatically determining a category label for each statement comprises assigning a category label for each statement when each statement first satisfies one of the rules.

12. The method of claim 2 wherein the rules include a catch all rule for categorizing statements that do not satisfy any of the other rules.

13. The method of claim 2 wherein the statements comprise a plurality of opening statements.

14. The method of claim 2 further comprising storing the rules.

15. The method of claim 2 further comprising collecting a plurality of statements to be categorized.

16. Software for the automated categorization of statements, the software-embodied in a computer-readable medium and operable to:

create one or more rules for categorizing the statements;

select one or more of the rules to apply to the statements;

apply the rules to a list of the statements; and

determine a category label for each statement based upon the rules.

17. The software of claim 16 wherein the statements comprise a plurality of opening statements.

18. The software of claim 16 further operable to create an output file, the output file including each statement and a corresponding category label.

19. The software of claim 18 wherein creating the output file comprises entering each statement and each corresponding category label into a spreadsheet.

20. The software of claim 16 wherein creating one or more rules comprises grouping the rules into one or more sets of rules.

21. The software of claim 16 further operable to display a graphical user interface.

22. The software of claim 16 wherein applying the rules to a list of the statements comprises applying the rules to the statements in a particular rule order in accordance with a rule hierarchy.

23. The software of claim 16 wherein applying the rules to a list of the statements comprises searching each statement for one or more text string combinations.

24. The software of claim 16 further operable to store the rules.

25. The software of claim 16 further operable to assign a category label to each statement when each statement first satisfies one of the rules.

26. A system for the automated categorization of statements, the system comprising:

a plurality of rules

a rule engine operable to create and store the rules used to categorize the statements; and

a category engine associated with the rule engine, the category engine operable to apply the rules to the statements and determine a category label for each statement.

27. The system of claim 26 wherein the statements comprise a plurality of opening statements.

28. The system of claim 26 further comprising a graphical user interface associated with the rule engine and the category engine, the graphical user interface operable to display the rules and the category labels.

29. The system of claim 26 wherein the category engine is further operable to create an output file.

30. The system of claim 29 wherein the output file includes each statement and a corresponding category label.

31. The system of claim 26 wherein the rule engine is further operable to group the rules into one or more sets of rules.

32. The system of claim 26 wherein the category engine searches each statement for one or more text string combinations to determine a category label for each statement.

33. The system of claim 26 wherein the rules include a catch-all rule for categorizing statements that do not satisfy any of the other rules.

34. The system of claim 26 wherein the category engine applies the rules to the statements in a particular rule order in accordance with a rule hierarchy.

* * * * *