

METHOD FOR IDENTIFYING AND CATEGORISING INFORMATION

FIELD OF INVENTION

The present invention relates to a method for identifying and categorising information from data for finance management services.

BACKGROUND OF THE INVENTION

Many personal finance software packages are available today to users. These packages range from basic Microsoft Excel sheets to the more evolved Microsoft Money or Quicken packages which may even generate graphs automatically. These packages allow users to analyse their income and expenditure to make better decisions on, for example, planning their finances to resolve debts or avoid getting into debts, as well as optimizing tax deductions, generating automatic tax reports, or even analyzing energy consumption and offering suggestions on applying for subsidies for better isolation in line with the Kyoto protocol. The data sources include and are not limited to bank statements, expenditure receipts and credit card statements of the users.

These packages are usually stand-alone programmes which run on a local personal computer. The data are either uploaded or input into the computer for subsequent analysis. For the convenience of their customers, financial institutions are enabling online access to data for example bank statements and credit card statements. Hence, there is a trend towards providing the existing software packages with a connection to the Internet to allow data from a variety of information sources accessible from the Internet to be retrieved and accepted for verification and categorisation.

Typical personal finance management software allows each localised user to import the bank statements into the software, either through seamless interface with the bank or by downloading the statements from online banking websites. The software will categorise the expenses based on the rules and knowledge compiled in the system based on previous interactions between the user and the software.

For example, each transaction listed in a bank statement has information pertaining to the merchant from which the transaction was held. From these information, an identifier is established automatically upon which the software decides how to categorise the transaction. If the identifier has been categorised previously, the system will simply categorise it in the previously determined category. If the identifier has not been previously categorised in the database, the system will request that the user determine the category. The system then stores the identifier in that category. This is known as tagging, hence, a new identifier is tagged in the user determined category.

The current software packages have the intelligence to provide a solution in categorising information such as expenses and income. The intelligence is built from the knowledge accumulated by the interaction with each localised user of the software. However, since the software packages run on a localised personal computer, the knowledge has to be built up by individual users, wasting time and resources.

Further, the knowledge accumulated is usually not exclusive to each user, having a centralised database for accumulating the knowledge for all users will save valuable time. Though this accumulated community knowledge is powerful, there may be unethical users who corrupt the knowledge database with redundant data. In such cases, the intelligence built up from the knowledge is likely to be corrupted or even destructive.

It is an object of the present invention to ameliorate or address one or all of the above problems.

SUMMARY OF THE INVENTION

According to an embodiment of the present invention, there is provided a method of identifying and categorising information from data which includes the steps of establishing an identifier from the data, identifying the identifier having a tagged category, initiating a first voting to select a preferred category for the identifier, providing a first category counter for the identifier in the preferred category, and increasing the count for the first preferred category counter of said identifier. If the identifier is identified

as not having a tagged category, the steps include tagging the identifier, providing a first category counter for the identifier in the tagged category, and initialising the counter.

The method further includes providing an option to initiate the first voting. When the option is not selected, the method will include the steps of triggering a system-operated selection.

Preferably, the method further includes generating a first report for the category tagged, voted or selected for the identifier, initiating a second voting and a second report for the updated category selected by the second voting.

In a second embodiment for multiple users, the method further includes providing a second category counter for the identifier in the preferred category of each user. The count in the first category counter is the sum of the counts in the second category counters of all users.

In a third embodiment, the method further provides a database which includes a plurality of identifiers, a category for each identifier and a first category counter for each identifier in the category, and matching the determined identifier to the plurality of identifiers in the database.

In a fourth embodiment for multiple users, the method provides a database which includes a plurality of identifiers, a category for each identifier and a first and a second category counters for each identifier in the category, and matching the determined identifier to the plurality of identifiers in the database.

BRIEF DESCRIPTION OF THE DRAWING

Figure 1 is a flowchart for the preferred embodiment of the invention.

Figure 2 is a flowchart for the second embodiment of the invention.

Figure 3 is a flowchart for the third embodiment of the invention.

Figure 4 is a flowchart for the fourth embodiment of the invention.

Figure 5 shows an architecture of the system where the invention is applied.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

A bank statement or credit card statement contains itemised data such as transaction information. The information in each item may include the location where the transaction was performed, the description of the transaction, the type of transaction – whether it is on credit or debit terms, the amount and date of the transaction. The identifier determined is usually a code or description in the transaction, such as the location of the transaction or the type of goods bought. The code or description is repeated if the transaction is repeated. For example if a first transaction was done at a location “Shopping Mall X”, then the identifier tagged to this location is determined as the identifier for this transaction. If a second transaction is done at the same location, “Shopping Mall X”, then a same identifier will be determined as the identifier again.

Figure 1 illustrates the flowchart of the preferred embodiment. For each transaction in the statement, the identifier is automatically established. The identifier is then identified if it has a tagged category as a result of a previous categorisation. When the identifier is identified as having a tagged category, a first voting is initiated to select a preferred category for the identifier. When the identifier is identified as not having a tagged category, the user will be asked to categorise the identifier, hence it will be tagged to a user-determined category, provided a first category counter for the identifier in the category and the counter is initialised.

Preferably, an option is provided to allow the system to select whether to initiate the first voting. When the system does not select the option, a system-operated selection will be triggered to select the category for the identifier.

The purpose of selecting, voting or tagging the identifier to a category is for consolidating and summarising transactions for analysis. The category can be general such as clothes, grocery, entertainment, sports, and transport, or detailed such as shirt, pants, wine, vegetables, petrol, movie and air tickets.

Further, a first category counter is provided for the identifier in the preferred category to count the number of times the preferred category is tagged to the identifier. The counter is initialised when an identifier is tagged to the category for the first time. There is an increment to the preferred category counter for the first and each subsequent time the identifier is assigned to be in that preferred category. The count in the first category counter is then increased each time the category is assigned.

In the system-operated selection, the first category counters for the identifiers in the various categories are compared. For example, in an item of a bank statement, an identifier "Shopping Mall X" is established. This identifier has been tagged to the categories "Clothes", C_X , "Grocery", G_X , and "Sports", S_X , previously. In this example, the first category counters for the identifier in the categories are C_X , G_X , S_X with counts 2, 12, 14 respectively. The last increment for C_X was made on 1st June 2006, for G_X was on 15th May 2006 and for S_X was on 30th April 2006. The system then selects the preferred category by any one of the following criteria:

- (a) category having the most recent increment from the first category counter, thus being the most recent preferred category for the identifier;
- (b) the first category counter having the highest count, or highest frequency of assignment;
- (c) a combination of both factors with a varying weighting; or
- (d) any other algorithm taking more complex analysis into account, such as Frequency analysis (Fourier transforms) of the voting process, second order moments (variance) of the voting distribution or even higher order moments.

For a selection based on criteria (a), "Clothes" will be chosen as the preferred category as "Clothes" was last tagged to the preferred category on 1st June 2006. Based on criteria (b), "Sports" will be chosen as "Sports" was tagged to the preferred category 14 times.

Based on criteria (c), according to the date of the last increment, the categories are assigned hypothetical weightings according to the following table:-

Category	"Clothes"	"Grocery"	"Sports"
Weighting _{date}	6	4	2

According to the date of the last increment in the counters, the categories are assigned hypothetical weightings according to the following table :-

Category	"Clothes"	"Grocery"	"Sports"
Weighting _{count}	1	6	7

Hence, after summing the weightings as shown below, the category "Grocery" will be selected, having the highest total weighting points.

Category	"Clothes"	"Grocery"	"Sports"
Weighting _{date}	6	4	2
Weighting _{count}	1	6	7
<u>Total</u>	<u>7</u>	<u>10</u>	<u>9</u>

The above steps are repeated for all the transactions in the statement. When all the transactions are processed, preferably, a first report is generated, displaying the user-determined or system-selected category for each identifier based on any of the criteria previously mentioned. This report allows the user to review the categories assigned to decide whether there is a need for a re-assignment. Advantageously, in case where there is any corrupted data causing undesirable assignment of categories, the user is alerted by reviewing this report.

The preferred embodiment further includes initiating a second voting to allow a user to re-assign the categories selected by the system or first determined by the user. In the second voting, a list of categories is provided for the user to select a preferred category for the identifier. The list of categories includes any categories previously defined or used.

Further, an alternative or additional option is provided in the user-operated voting for the user to define a new preferred category for the identifier.

For example, an identifier is "Shopping Mall X". In one transaction, a user may have bought food from "Shopping Mall X", so a category "Grocery", G_x , is selected. In another transaction, the user may have bought clothes. Hence, initiating the user-operated voting allows the category of the identifier in the latter transaction to be overwritten as "Clothes", C_x , this being the preferred category. If the category "Clothes" has not been defined previously, the user has the option to define this new category which will then be incorporated to the list of categories.

When an identifier is voted into a category which has not been previously tagged to the identifier, a first category counter will be provided for the identifier and the counter is initialised. If the category has been previously tagged to the identifier, an option is provided to decide whether to increase the count in the first category counter or not to increase the count if it has already been increased previously prior to generating the first report.

After the categories selected, voted or tagged to all the identifiers have been finalised, preferably a second report indicating the preferred category for the identifier is generated. Further, the second report may further consolidate and summarise the transactions by categories, thus providing an analysis of the finances of a user. This report is useful in helping the user to plan his finances to resolve debts or avoid getting into debts, among other possibilities as discussed in paragraph one of the section 'Background of the Invention'.

A second embodiment of this invention is applicable for multiple users as shown in Figure 2. In this embodiment, the method includes the steps of establishing an identifier from the data and identifying the identifier having a tagged category. When identifying the identifier having a tagged category, the categories used by other users are taken into consideration. In this embodiment, the system would be linked to other systems via a local area network, intranet or the Internet.

For the identifier identified as having a tagged category, a first voting is initiated for a user to select a preferred category for the identifier, providing a first category counter for the identifier in the preferred category, providing a second category counter for the identifier in the preferred category of each user, and increasing the count for the second category counter of said identifier. The count in the first category counter is the sum of the counts in the second category counters of all users.

For the identifier identified as not having a tagged category, the steps include tagging the identifier to a user-determined category, providing a first and a second category counters for the identifier in the tagged category, and initialising the counters.

An option is provided to allow the system to decide whether to initiate the first voting. When the first voting is not initiated, a system-operated selection will be triggered to

select a category for the identifier, based on the first category counters for the identifier and/or the second category counters for the identifier.

Using an example for illustration, users P, Q, R, S have the same identifier "Shopping Mall X" in their transactions. Users P and R have assigned the identifier to the category "Grocery", while Q and S have assigned the identifier to the category "Clothes".

The first category counter for the identifier "Shopping Mall X" in the category "Grocery" is G_X . The second category counter for the identifier "Shopping Mall X" in the category "Grocery" for user P is G_{XP} , and for user R is G_{XR} . The first category counter G_X is the sum of the second category counters, G_{XP} and G_{XR} , for the identifier, therefore, $G_X = G_{XP} + G_{XR}$. Similarly, the first category counter for the identifier "Shopping Mall X" in the category "Clothes" is C_X , the second category counter for the identifier "Shopping Mall X" in the category "Clothes" for user Q is C_{XQ} , and for user S is C_{XS} . Similarly, $C_X = C_{XQ} + C_{XS}$.

The provision of the second category counters is to allow the system operated selection another option to take into consideration all the categories used by other users when selecting the preferred category. This is to enable the utilisation of the accumulated community knowledge. For example, in addition to the criteria available in the preferred embodiment for the system to select a preferred category, the criteria for selecting may now further include:

- (e) category having the most recent increment from the second category counter, thus being the most recent preferred category for the identifier;
- (f) the second category counter having the highest count, or highest frequency of assignment;
- (g) a combination of any of the factors with a varying weighting; or
- (h) any other algorithm taking more complex analysis into account, such as Frequency analysis (Fourier transforms) of the voting process, second order moments (variance) of the voting distribution or even higher order moments.

Preferably, the method further includes generating a first report for each user. In the report, the category selected by the system or tagged or voted by the user is displayed for the user to review and determine if there is a need for re-assignment of the categories. If the user wishes to re-assign the categories, a second voting is initiated. When an identifier is voted into a category which has not been previously tagged to the

identifier, a first and a second category counters will be provided for the identifier and the second category counter is initialised. If the category has been previously tagged to the identifier, an option is provided to decide whether to increase the count in the second category counter or not to increase the count if it has already been increased previously prior to generating the first report.

After the categories assigned, voted or tagged to all the identifiers have been finalised, preferably a second report is generated. This second report displays the preferred category for the identifier, and further consolidates and summarises the transactions by categories. From the consolidation and summary, the user is able to monitor his finances for budgeting or any other purposes.

In a third embodiment of the invention as shown in Figure 3, a method of identifying and categorising information from data includes the steps of establishing an identifier from the data, having a database which includes a plurality of identifiers, a category for each identifier and a first category counter for each identifier in the category, identifying the identifier having a tagged category in the database by matching the identifier to the plurality of identifiers, optionally initiating a first voting to select a preferred category for the identifier, and increasing the count for the first preferred category counter of said identifier. When the identifier is identified as not having a tagged category, the steps include tagging the identifier to a user-determined category, providing a first category counter for the identifier in the category of a user. The first category counter is initialised and the count is updated into the database.

Preferably, an option is provided to decide whether to initiate the first voting. When the first voted is not initiated, the identifier will be tagged to a category by a system operated selection based on the first category counters for the identifier.

In a fourth embodiment for multiple users as shown in Figure 4, a method of identifying and categorising information from data includes the steps of establishing an identifier from the data, having a database which includes a plurality of identifiers, a category for each identifier, a first and a second category counters for each identifier in the category, identifying the identifier having a tagged category in the database by matching to the plurality of identifiers, initiating a voting to select a preferred category for the identifier, and increasing the count for the first preferred category counter of said identifier. When the identifier is identified as not

having a tagged category, the steps include tagging the identifier to a user-determined category, providing a first and a second category counters for the identifier in the category of a user. The first category counter calculates the sum of all the second category counters for all users. The category counters are initialised and the results are updated into the database.

Preferably, an option is provided to decide whether to initiate the first voting. When the first voting is not initialised, a system-operated selection will select a category for the identifier, based on the first and second category counters of the identifier.

Having a database for the identifiers, the categories previously defined and the respective first and second category counter eases the management of the community knowledge. Further reports on the identifiers available, the categorisation and statistics available can be generated from the database for any purpose such as data analysis.

The other steps for the third and fourth embodiments found in the flowchart are similar to those discussed for the preferred and second embodiments respectively.

Figure 5 illustrates how the invention can be implemented on a computer system. Data such as transactional information from bank statements stored on computer servers of the banks. Similarly, identifiers such as location, description or type of transaction are stored on computer servers or any databases. These computer servers and databases are accessible by a network, an Intranet or an Internet connection. The figure shows how the categorisation application can be located on a standalone server accessible using a network. Optionally, the categorisation engine can be software running on the PC clients of the users, which would result in a distributed categorisation engine and a centralised database to store the tagging between the identifiers and the categories.

The invention provides a method of identifying an identifier and categorising information from data, such as bank statements which are accessible from the computer server of a bank. From the information in the data, an identifier is identified, the identifier is determined to have a tagged category from the source such as a listing or database of identifiers and categories. For an identifier having a tagged category, an option is provided for the system to initiate a first voting to select a preferred category for the identifier. A first category counter is provided for the identifier in the preferred category. For an identifier not having a tagged category, a user determined category is tagged to the identifier and a first category counter is provided

and initialised for the identifier in the category. Thus, the knowledge of categorising is shared between multiple users.

We claim:

1. A method of identifying and categorising information from data for a user which includes the steps of:

- (a) establishing an identifier from said data;
 - (b) identifying said identifier having a tagged category;
 - (c1) initiating a first voting to select a preferred category for said identifier;
 - (c2) providing a first category counter for said identifier in said preferred category;
 - (c3) increasing the count for said preferred first category counter of said identifier;
- wherein when said identifier is identified as not having a tagged category, the steps

include

- (d1) tagging a user determined category to said identifier;
- (d2) providing a first category counter for said identifier in said category;
- (d3) initialising said first category counter.

2. A method as claimed in claim 1 wherein prior to the step (c1), the method further includes the step of

- (c) providing an option to initiate said first voting.

3. A method as claimed in claim 2 wherein said first voting is not initiated, further includes the step of triggering a system operated selection .

4. A method as claimed in any of the preceding claims 1 to 3, further includes the step of

- (e) generating a first report,

wherein said identifier with said category tagged, voted or selected is shown.

5. A method as claimed in any of the preceding claims 1 to 4, further includes the step of

- (f) initiating a second voting.

6. A method as claimed in claim 5 wherein said second voting further includes the step of

- (g) providing a list of categories for a user to enter a preferred category for said identifier.

7. A method as claimed in the preceding claim 5 or 6 wherein said second voting further includes the step of

(h1) providing a new category for said user to enter as the preferred category for said identifier.

8. A method as claimed in claim 7, further includes the step of

- (h2) providing a first category counter for said identifier in said preferred category;
- and
- (h3) initialising said first preferred category counter.

9. A method as claimed in claim 8 further includes the step of

- (i) generating a second report indicating said preferred category for said identifier.

10. A method of identifying and categorising information from data for a plurality of users which includes the steps of:

- (a) establishing an identifier from said data for each user;
 - (b) identifying said identifier having a tagged category;
 - (c1) initiating a first voting to select a preferred category for said identifier;
 - (c2) providing a first category counter for said identifier in said preferred category;
 - (c3) providing a second category counter for said identifier in said preferred category of each user;
 - (c4) increasing the count for said preferred first category counter of said identifier;
- wherein when said identifier is identified as not having a tagged category; the steps

include

- (d1) tagging a user-determined category to said identifier;
- (d2) providing a first category counter for said identifier in said category;
- (d3) providing a second category counter for said identifier in said category;
- (d4) initialising said second category counter;

wherein the count in the first category counter is the sum of the counts in said second category counters for all users.

11. A method as claimed in claim 10 wherein prior to the step (c1), the method further includes the step of

- (c) providing an option to initiate said first voting.

12. A method as claimed in claim 11 wherein said first voting is not initiated, further includes the step of triggering a system operated selection.

13. A method as claimed in any of the preceding claims 10 to 12 further includes the step of
(e) generating a first report,
wherein said identifier with said category tagged, voted or selected is shown.
14. A method as claimed in any of the preceding claims 10 to 13, further includes the step of
(f) initiating a second voting.
15. A method as claimed in claim 14 wherein said second voting further includes the step of
(g) providing a list of categories for a user to enter a preferred category for said identifier.
16. A method as claimed in the preceding claims 14 or 15 wherein said user-operated voting further includes the step of
(h1) providing a new category for said user to enter as the preferred category for said identifier.
17. A method as claimed in claims 16, further includes the step of
(h2) providing a first category counter for said identifier in said preferred category;
and
(h3) providing a second preferred category counter for said identifier in said preferred category of each user.
18. A method as claimed in claims 14 to 17 further includes the step of
(i) generating a second report indicating said preferred category for said identifier.
19. A method of identifying and categorising information from data which includes the steps of:
- (a) establishing an identifier from said data;
 - (b) having a database which includes
 - i. a plurality of identifiers;
 - ii. a category for each identifier;
 - iii. a first category counter for each identifier in said category;
 - (c) identifying said identifier having a tagged category in said database by matching to said plurality of identifiers;
 - (d1) initiating a first voting to select a preferred category for said identifier;
 - (d2) initialising said first preferred category counter; and

(d3) increasing the count for said preferred first category counter of said identifier;
wherein when said identifier is identified as not having a tagged category in said database, the steps further include

- (e1) tagging a user-determined category to said identifier.
- (e2) providing a first category counter for said identifier in said category;
- (e3) initialising said first category counter; and
- (e4) updating said first category counter into said database.

20. A method as claimed in claim 19 wherein prior to the step (d1), the method further includes the step of

- (d) providing an option to initiate said first voting.

21. A method as claimed in any of the claims 19 or 20 wherein said first voting is not initiated, further includes the step of triggering a system-operated selection.

22. A method as claimed in any of the preceding claims 19 to 21 further includes the step of
(f) generating a first report,
wherein said identifier with said category tagged, voted or selected is shown.

23. A method as claimed in any of the preceding claims 19 to 22, further includes the step of
(g) initiating a second voting.

24. A method as claimed in claim 23 wherein said second voting further includes the step of
(h) providing a list of categories for a user to enter a preferred category for said identifier.

25. A method as claimed in the preceding claims 23 or 24 wherein said user-operated voting further includes the step of

- (i1) providing a new category for said user to enter as the preferred category for said identifier;
- (i2) providing a first category counter for said identifier in said preferred category; and
- (i2) updating said new category and said first preferred category counter for said identifier into said database.

26. A method as claimed in claims 23 to 25 further includes the step of

- (j) generating a second report indicating said preferred category for said identifier.

27. A method of identifying and categorising information from data for a plurality of users which includes the steps of:

- (a) establishing an identifier from said data for each user;
- (b) having a database which includes
 - i. a plurality of identifiers;
 - ii. a category for each identifier;
 - iii. a first category counter for each identifier in said category;
 - iv. a second category counter for each identifier in said category of each user,wherein said first category counter is the sum of said second category counter for all users;

(c) identifying said identifier having a tagged category in said database by matching to said plurality of identifiers;

(d1) initiating a first voting to select a preferred category for said identifier;

(d2) increasing the count for said preferred second category counter of said identifier; wherein when said identifier is identified as not having a tagged category in said

database, the steps further include

- (e1) tagging a user-determined category to said identifier.
- (e2) providing a first category counter for said identifier in said category;
- (e3) providing a second category counter for said identifier in said category;
- (e4) initialising said first and second category counters; and
- (e5) updating said first and second category counters into said database.

28. A method as claimed in claim 27 wherein prior to the step prior to (d1), the method further includes the step of

- (d) providing an option to initiate said first voting.

29. A method as claimed in any of the preceding claims 27 or 28 wherein said first voting is not initiated, further includes the step of triggering a system-operated selection.

30. A method as claimed in any of the preceding claims 27 to 29 further includes the step of

- (f) generating a first report,

wherein said identifier with said category tagged, voted or selected is shown.

31. A method as claimed in any of the preceding claims 27 to 30 further includes the step of

- (g) initiating a second voting.

32. A method as claimed in claim 31 wherein said second voting further includes the step of
(h) providing a list of categories in said database for a user to enter a preferred category for said identifier.

33. A method as claimed in the preceding claims 31 or 32 wherein said user-operated voting further includes the step of

- (i1) providing a new category for said user to enter as the preferred category for said identifier;
- (i2) providing a first category counter for said identifier in said preferred category;
- (i3) providing a second preferred category counter for said identifier in said preferred category of each user;
- (i4) initialising said second preferred category counter; and
- (i5) updating said new category, said first and second preferred category counters for said identifier in said preferred category of each user into said database.

34. A method as claimed in claim 33 further includes the step of

- (j) generating a second report indicating said preferred category for said identifier.