3440 S. Jefferson St., #1125 Falls Church, VA 22041 202-607-8904 www.execware.com

Data Mining With PCs?

Well, not really, although the goal of a new methodology named contextual data modeling resembles that of data mining. Both processes are analytic tools that search databases to uncover information that is not obvious on a single flat table of data. Data mining involves massive databases probed by large computers, using mathematical formulas, to produce anaIyses—even forecasts — based on statistical theory.

Contextual data modeling involves small datasets, on personal computers, with the mind of the user defining and controlling the analysis. But their use is the same: decisions made by humans. However, that with contextual data modeling (CDM) the processing and decision-making are not separate. The user's mind and the data processing are interactive. CDM consists of a series of formative decisions, reaching answers to the initial question, and often answers to questions that arise spontaneously as the user studies that data from different perspectives. How does this happen?

Making decisions based on new knowledge is a philosophical matter. Immanuel Kant argued that empirical observation alone does not constitute knowledge. Kant said that knowledge requires two components: sensory perception plus a conceptual component of ordering and then interpreting. Here is an example. Assume you are driving, late in the day and tired, in a sparsely settled area. You glance at the gas gauge and perceive a low fuel level. Your mind associates that fact with your knowledge that your destination is 30 miles ahead, with your desire to arrive there that evening, with your feeling of fatigue, with your valuing alertness as essential when driving, with your belief that the next gas station may be far ahead, and with your recollection that you just left a town with gas stations and several motels. So, what do you decide to do?

Whether you continue ahead, or go back for fuel and then resume driving, or stay overnight and refuel in the morning results from your personal analysis of the situation. It started with a single datum: low fuel level. That was Kant's sensory perception. Your mind blended that new knowledge with your prior knowledge, goals, and values. Then your intuition and imagination kicked in, and you subconsciously prioritized (Kant's ordering) the many factors involved and reached a final interpretation of what they alI—not just the single datum—meant you should do.

You also look at an airport monitor seeking a single datum: the correct aircraft. But that same data table contains a wealth of information needed by other peopIe—those interested in the data for other reasons. Control tower staff want to see the table with times in the first column. Persons handling gate assignments want gate numbers in the first column. A person curious about traffic to a particular city wants destinations in the first column. And in each of those cases, different arrangements of the remaining columns provide different information. How, then, can a person extract from a database all the useful information it contains? One answer is data mining. Another answer is CDM. Both terms use the same word, modeling, but the meaning differs.

Data mining starts with a model of what is intended, as with a business model for a new

venture, or a model of a boat one plans to build. Algorithms are designed accordingly, and the data are processed to derive the answer generated according to the algorithms. Modeling with CDM, however, is similar to modeling a lump of clay to change it into a useful object. And as we do that we look at it from many different perspectives. With CDM we do more than read outputs–we add motion/ movement as a key factor in the analysis of a data table. How do we do that?

A data table can be modeled several ways. First, the columns can be rearranged. Second the table can be shortened or lengthened by re-selecting the items to be listed, or dragging the bottom border down or up. Third, the values of the parameters of items on the table can be revised. Why? Thousands of different tables can be designed, each giving the decision-maker new knowledge. But that takes time. And what a person sees is held for processing primarily in short-term memory. So potentially valuable new knowledge can fade quickly while we wait for a new printout. Further, the pattern of our current reasoning may be lost unless the desired changes are made immediately. (We don't understand a statue by looking at the front, then a side minutes or hours later, etcetera). Consequently, the table user must be able to easily and interactively introduce changes, immediately see the results, and continue that process until all possible useful information has been disclosed. Data mining does not lend itself to that. With CDM, the analyst can see—in a controlled series of varied tabIes—the values of selected parameters, and get their relationships in the context of (1) all the displayed parameter values, (2) other data about the items on the table, and (3) his or her current line of reasoning. Thus the name contextual data modeling.

Our goal was giving to analysts of small datasets the ability to personally interact with their data using only three primary screens: Query (design a data table), Table (see the query result), and a virtual stack of the dialog boxes, i.e. Item Specifications used to enter details of an item into the computer. Editing parameter values causes the table to be automatically re-sorted—exceIIent for "what if"' questions. That required one simple invention: combining on one screen the controls for the select and sort functions. A patent was granted.

Then, to better manage examining all possible permutations of a data table (for example, 6 columns can be arranged 720 ways) we invented automated sorting. It is not logical to consider a permutation useless until it is examined. Another patent was granted. But looking at 720 tables is almost prohibitive. Yet the common practice of redesigning the table to have fewer columns means you won't see some data. So we patented keeping all columns showing but marking some for exclusion from sorting.

Consistent with the goal of shortening the link between the user's mind and the monitor, a CDM program contains, on each Item Specification, lengthy descriptions. Also listed are links to other items in the dataset, not just those on the table, that relate to the one being viewed. And to that Item Spec the user can attach notes containing temporary remarks. Then, with the Item screen open, the user flips back and forth through similar screens for all items on the current table. Sensory perception is enriched in the process.

Comparing a CDM program with other software is like comparing oranges with tomatoes, both being fruit. Our Reason® , the only software incorporating our patented technology, is a new and different type of software that fits in no existing category.

Robert Listou