VISUAL DATA EXPLORATION (VDX)

for Data, Migration and Scenario Analyses

The IT landscapes of the financial services and public administration sectors are marked by large amounts of data having multiple relations. Because the data normally has evolved over time, assuring data quality can be problematic. Analyzing and testing mass data during migration, for example, requires extra effort and money. The same holds true for scenario analyses: e.g., proposed changes to premiums or other policy changes frequently present particular challenges when it comes to testing the scenarios, since business-critical cases must be pinpointed and processed on an individual basis, and under severe time constraints.

It's a fact that conventional testing methods focus on program quality, and not on the quality of the data itself. As a result, their effectiveness is limited in cases where relatively simple programs, like those used for data migrations, must be applied to very complex data pools.

A particular problem with conventional scenario analyses is the frequent lack of reference values – or, at least, the latter can be difficult to generate. The primary aim of testing thus becomes the prevention of gross deviations within individual datasets. However, using tabular display methods, such discrepancies are extremely difficult to pinpoint, even for individuals with a thorough understanding of the mapped relations. From an image standpoint, problematic datasets absolutely need to be dealt with, yet they frequently deviate only to within tenths of a percentage point.

Visual data exploration offers a highly effective alternative solution for testing large-data applications. The use of statistical techniques permits data structures and patterns to be graphically represented, thus enabling existing deviations from reference values to be identified within a fraction of the time that it takes using conventional means.

Visual data exploration goes beyond random sampling or testing against reference values, however, as it actually permits data pools to be analyzed fully: The analyzer is able to identify all deviations down to the individual dataset. Tools capable of graphically representing large data pools via explorative-descriptive statistical methods are available to support the task. The key to success is a well trained eye for reviewing the data and for selecting the proper means of preparing it graphically. In addition, such factors as data heterogeneity, missing attributes or additional characteristics can be integrated into the final evaluation.

Ultimately, this innovative approach represents an entirely new field of activity for test specialists, since it places the focus on identifying deviating datasets as opposed to program errors. The advantages are multifold:

- Visual data exploration (VDX) can be implemented with considerably less time and effort
- Used in combination with conventional testing methods the technique delivers extremely reliable results
- VDX makes concrete information based on individual datasets available to decision makers
- VDX helps save costs – due to the comparable ease and speed of implementation
- VDX can be applied early in the development cycle to help identify problems earlier, for more effective error prevention

The Procedure

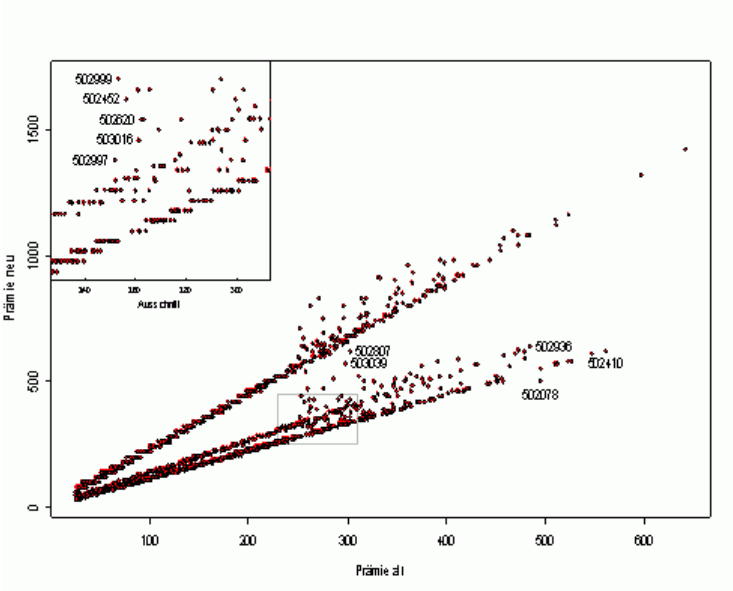
The OBJENTIS approach to visual data exploration can best be defined as a procedure for interpreting the condition and processability of large, highly dimensional datasets.

During the development phase for implementing VDX, the analyzer must work closely with the expert business users and also have a thorough understanding of the planned changes. In addition, in-depth know-how of software development and statistical methods is required to be able to analyze the changes, select the suitable graphic visualization methods and identify any “problematic” data relations. With respect to graphics, they will yield more powerful information when tailored to the specific business case, the selection of which also requires a well-founded understanding of the business (user) application. Available statistical tools include open source solutions (e.g., “R suite”).

The ongoing execution of tests can be performed by trained internal testers, with the core skill required for VDX being the ability to interpret the graphics. Even if massive amounts of data are involved, the visualization methods used enable the tester to identify individual dataset deviations within minutes. Such deviations point to potential program errors, which means the tester can immediately evaluate the situation and react proactively.

What’s more, external testing know-how and the active cooperation of the company’s business experts are only required during the initial VDX set-up phase, and not during ongoing operations.

Sample Application: Migrating Insurance Premium Data



The graph above reveals the following information for a sample migration of insurance premium data:

A portion of the premiums remained virtually the same (lower curve) after migration, another portion (middle curve) increased slightly, and the rest (top curve) doubled because annual, as opposed to semi-annual, premiums were considered. The detailed insert clearly shows that, in calculating the new premiums, the values have been rounded off (“stepped”).

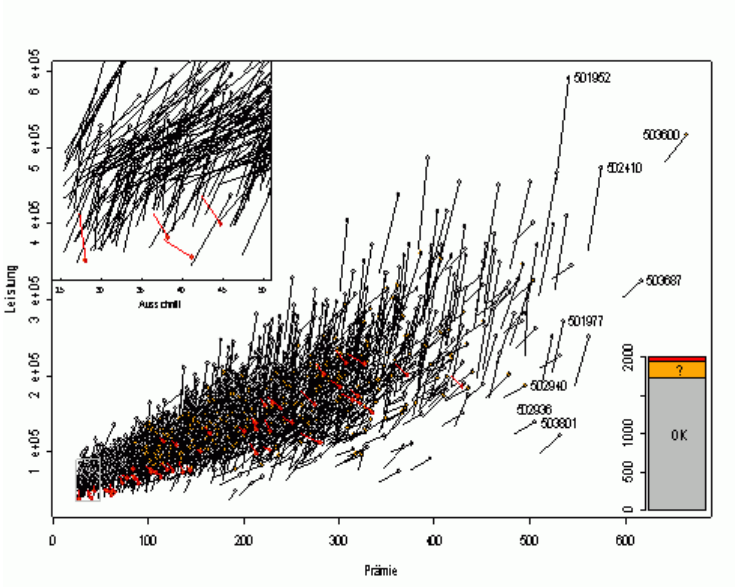
Also obvious from the graph, at a premium value of 250 and above we see plots that deviate from the respective curves – any of which could potentially be due to a program error. The tester would next select several of the deviating plots and assign each the corresponding policy number for identification purposes. In addition, the deviant datasets are automatically saved to a CSV file for further testing.

Because the deviant datasets can be individually selected and processed further, an adaptation of this sort is significantly more efficient than trying to evaluate deviations using the migration program itself.

Sample Application: Product Adaptation

Assume that an insurance company wants to develop a new product and that this requires re-calculating policies for over 100,000 clients, ranging from private customers to large institutions. The challenge is to analyze whether or not the new product offering will be financially attractive for all customers, and to what extent it will be profitable for the insurance company. The large number of datasets makes an analysis by the company's business experts (stakeholders) difficult, at a high cost in time and money.

A graphical analysis is the best solution, as it can be used to identify all cases that do not meet the targets – e.g., cases where a change results in a lower benefit at a higher premium, or where the new premium does not stand in relation to the increase in benefit, etc. – without overly burdening company personnel. Using graphics, such cases can therefore be accurately identified and tested at a fraction of the usual time and cost.



Specifically, the existing tariff scheme needs to be replaced by a new one, with both the premium and benefit for each policy targeted for an increase. The needle diagram above not only provides vital feedback to decision makers for use in developing the new tariff scheme, but it can also be applied for testing during the coding phase.

The point of each needle marks the premium and benefit amount for a given policy prior to changing to the new tariff, while the head represents the new values after the change. Most of the needles “tilt” to the upper right (from point to head), indicating a significant increase in both the premium and benefit (also see the detailed view).

Some needles, however, do not fit the pattern, and therefore indicate potential errors. These are identified automatically and are marked in red. In addition, the heads of those needles with a premium change of more than 20 Euros are marked in orange. This gives the tester a quick yet reliable overview of the total number of datasets exhibiting significant premium changes, including the associated data distribution.

Potential Uses

In addition to providing testers with an excellent overview of the data structure, VDX is ideal for:

- Data pool, migration and scenario analyses in large-data applications
- Performing marginal analyses
- Error prevention, by helping to identify errors during the project's earliest phases
- Test data engineering; defining reference values
- Test case identification and explorative testing

VDX provides invaluable decision making support throughout the entire development and testing process, from requirements engineering through to acceptance testing, and is ideally suited for use in insurance, banking and public administration applications.

OBJENTIS understands what it takes to help customers get the most out of visual data exploration. Not only can we help you develop and implement the technique for your purposes, but we will also train your organization's test experts in the supporting analysis methods.

Contact:

OBJENTIS Software Integration GmbH
Mariahilfer Strasse 3, A - 1060 Vienna, Austria
Tel +43/1/585 68 00-0
Fax +43/1/585 68 00-50
e-Mail info@objentis.com