

Ph.D. Lecture

In partial fulfillment of the terms for obtaining the Ph.D.-degree, Morten Middelfart will give a lecture on the following subject:

Sentinel Mining

on Monday 6th of December 2010, 13.00, in room 02.12 at Selma Lagerlöfs Vej 300

Abstract:

This thesis introduces the novel concept of sentinel rules (sentinels). Sentinels are intended to represent the relationships between the data originating from the external environment and the data representing the critical organizational performance. The intention with sentinels is to warn business users about potential changes to Key Performance Indicators (KPIs) and thereby facilitate corrective action before such a change becomes a reality.

Specifically, sentinels are rule relationships at the schema level in a multidimensional data cube. These relationships represent changes over time in certain measures that are followed by a change in a user defined critical measure, typically a KPI. An important property of a sentinel is bi-directionality, which means that the change relationship holds in the complement direction, since a sentinel with the bi-directional property has a higher chance of being causal rather than coincidental. Sentinels can vary in complexity depending on the number of measures that are included in the rule: Regular sentinels represent relationships where changes in one measure lead to changes in another within a given time frame. Generalized sentinels represent relationships between changes in multiple measures leading to changes in a given measure within a given time frame. Multidimensional sentinels combine the schema and the data levels, meaning that each measure change in the rule can hold for either subsets or the entire cube. A generalized sentinel could for example notify users that revenue might drop within two months if an increase in customer problems combined with a decrease in website traffic is observed, whereas a multidimensional sentinel could warn users that revenue might drop within two months if an increase in customer complaints in USA (drilldown into geography dimension) combined with a decrease in the money invested in customer support for laptop computers (drilldown into product dimension) is observed.

The work leading to this thesis progressed from algorithms for regular sentinel mining with only one source and one target measure, into algorithms for mining generalized and multidimensional sentinels with multiple source measures. Furthermore, the mining algorithms became capable of automatically fitting the best warning periods for a given sentinel. Aside from expanding the capabilities of the algorithms, the work demonstrates a significant progression in the efficiency of sentinel mining, where the latest bitmap-based algorithms, that also take advantage of modern CPUs, are 3–4 orders of magnitude faster than the first SQL-based sentinel mining algorithm. This work also led to the industrial implementation of sentinel mining in the commercial software TARGIT BI Suite, which attracted the attention of leading industry analysts. In short, the work in this thesis has turned sentinel mining from a theoretical idea into concrete, highly efficient algorithms, and in addition it has demonstrated sentinels to be useful and unique.

All interested parties are welcome. After the defense the department will be hosting a small reception in cluster 3.