Facharbeit.

Topic: Visualization of nutrient/nutrient relationships (correlations)

Problem description: The Swiss feed database contains data about nutrient contents of animal feed. Some data sets, such as the hay survey, are large enough to be displayed as color spots on a google map together with a scatter chart showing the selected nutrient content over a time axis (figure 1). The temporal scatter chart is a useful tool to visualize the extent of variations, yearly fluctuations and possible time trends. Currently up to three nutrients can be displayed at a time (figure 2). However, questions such as *how do nutrients correlate with the enregy content* or *how do nutrients correlate with each other*, cannot be derived from the current result view.

Figure 1. Screen shot of the result view of a query about the hay survey displaying the energy content (2005-2011)
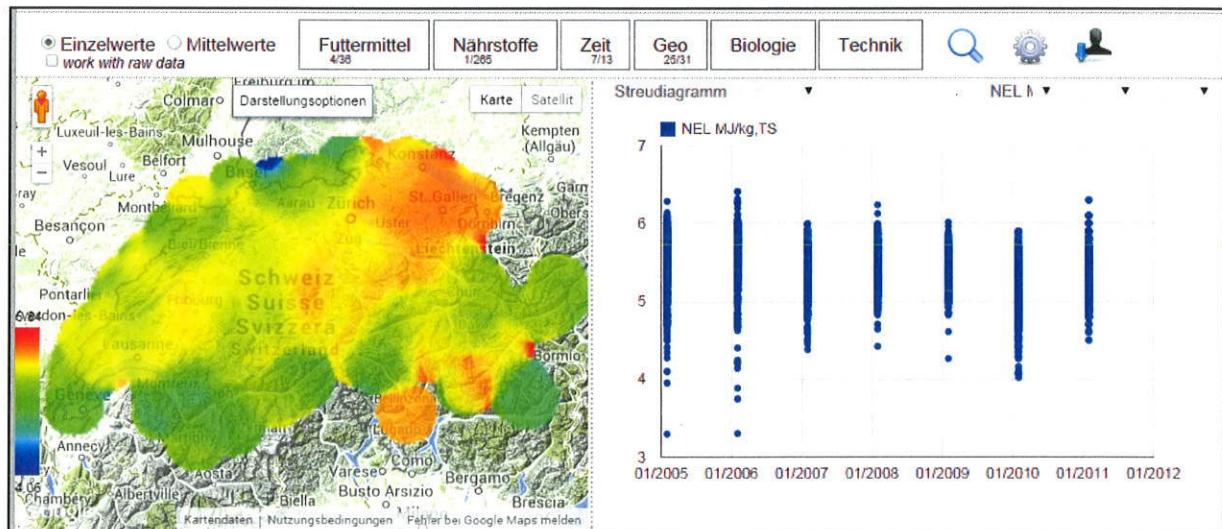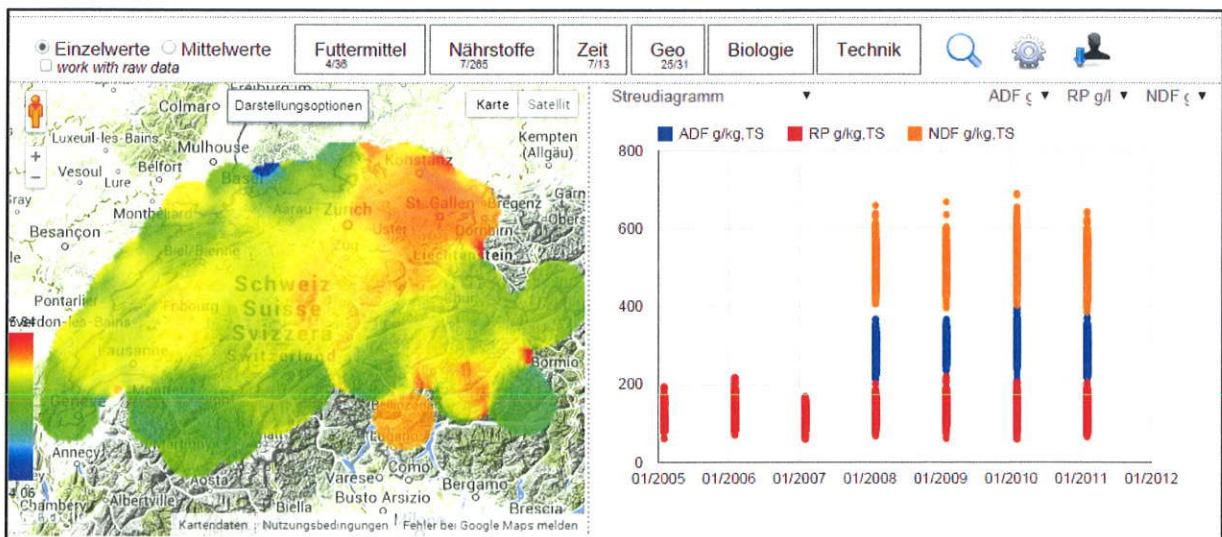

Figure 2. Screen shot of the result view of a query about of the hay survey displaying the nutrients RP, ADF, NDF (2005-2011)

The goal of the project is to implement the new functionality *correlated nutrients* that contains the following elements:

- For the selected nutrient(s) a user can specify which nutrient is considered independent and represented on the x-axis and which nutrient(s) are defined as dependent parameters and displayed on the y-axis.
- The chart type is a scatter chart. Each nutrient has a distinctly different colour or shape. The legend of the x-axis must be shown.
- By default, the result view is restricted to 150 saples which the user can manually modify under the *advanced option* button. The implemented functionality must be optimzed for fast response time. Particularly it should be tested whether the default setting must be maintained or not. The defailt setting could lead to misinrterpretions if not the full sample number is considered. In case of a restricted result view, users should be informed by a comment line or window pop-up.
- Minimize the number of points displayed in the scatter chart in order to make the application faster
- If a good correlation exists between two nutrients, derived regression equrations can be used for estimation purposes or data quality control (see figure 3 as example). As part of the project it should be tested whether the implementation of calculated correlations and regressions would significantly slow down the system performance.

Figure 3. Scatter chart of correlated nutrients including coefficient of determination (R^2 = squared correlation) and regression line.