

FMDB: Activities, Plans and Action Points

Andrej Taliun

January 18, 2012

Agenda

- Overview of Activities (Andrej Taliun)
- A temporal model of the FDB (Samuele Zoppi)
- Use and functionality of FDB (Annelies Bracher)
- Demo of FDW (Francesco Cafagna)
- Description and processing of feed data (Annelies Bracher)
- Plans and Action Points (Andrej Taliun)

History of the Feed Database

- Started by ETH and Agroscope in 2005:
 - collection, cleaning and structuring the feed data,
 - non-temporal design and implementation of the database,
 - user interface with build in queries,
 - public service at www.feed-alp.admin.ch.
- UZH replaces ETH in June 2010:
 - August - the database is run by UZH,
 - October - SNF proposal that establish collaboration plan for the next 3 years is submitted,
 - November - last technical issues related to migration are resolved,
 - June/November - frequent meetings at least once per month.

Current State of the Feed Database

■ Setup:

- run on a virtual machine with 2Gb of RAM and 2.3 GHz CPU,
- Apache web server,
- Implemented with PHP,
- MySQL database management system.

■ Maintenance:

- Andrej Taliun - administration: identifications of problems and corrections, management of the database;
- Annelies Bracher - communication with users, all aspects of data collection, cleaning and import;
- Technicians at UZH - software and hardware setup.

Unresolved Issues After Migration

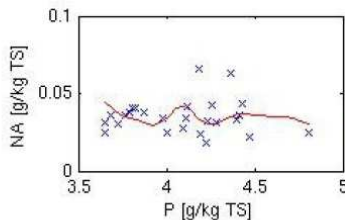
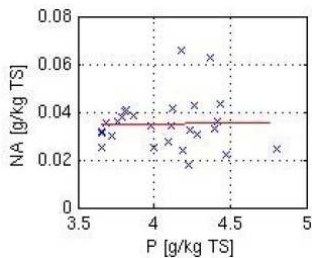
- We do not have complete understanding on the code:
 - interface design, functional and data requirements are mixed,
 - hacks are present,
 - almost no documentation,
- The chosen system architecture is not easy extendable:
 - limited support for math, file formats and security,
 - no support for object oriented programming and team working.

Current Development of the Feed Database

- Student Project: Online Statistical Computation in the Feed database
 - Duration: October 15 - March 15,
 - Student: Zafer Adigüzel,
 - Goal: comparison of linear and kernel regression methods to identify correlation between nutrients,
 - Data: measurements of nutrients in Barley for 1992-2009,
 - Methodology:
 - Graphical and numerical comparison of methods for various periods and pairs of nutrients,
 - Experimental evaluation on imputing missing measurements.

Current Development of the Feed Database

- Student Project: Online Statistical Computation in the Feed database
 - linear regression assumes that the data is linear,
 - kernel method does not place any assumptions on data distribution.



Current Development of the Feed Database

- Student Project: Online Statistical Computation in the Feed database
 - Key Results: linear regression results on a smaller error for the feed data.

<i>Missing values for the nutrient</i>	<i>Basis for regression is the nutrient</i>	<i>Year of analysis</i>	<i>Error (linear regression)</i>	<i>Error (kernel regression)</i>
TSL	RA	Whole given period	14.1645	15.7086
TSO	TSL	Whole given period	17.1067	19.5016
RP	RA	Whole given period	8.4512	8.4627
NA	P	Whole given period	0.0584	0.0592

Current Development of the Feed Database

- Student Project: Temporal Model of the Feed Database
 - Samuele Zoppi, finished at February 15,
 - Yannick Widmer, started in August,
- Goal: design of a temporal model for the feed database,
- Data: measurements of nutrients in Barley for 1992-2009,
- Methodology:
 - deriving current data requirement and database design,
 - refinement of the database design to support single measurements,
 - implementation of the model and experimental validation.

Current Development of the Feed Database

- Temporal Model of the Feed Database (Samuele Zoppi and Andrej Taliun)
 - preliminary temporal model is ready and validated,
 - postgresql database management system is installed,
 - two database repositories are created: clean and for testing,
 - currently, access is granted for internal use only.

Current Development of the Feed Database

- The Data Warehouse (Francesco Cafagna):
 - design supports historical and atomic information,
 - preliminary version is implemented,
 - filled with all currently available data,
 - validated for detailed and (partially) aggregated queries.

Goals for 2011

- Temporal database with all available history,
- On-line computation of up-to-date summaries,
- User interface will all current functionality,
- User interface to extract historical information,
- Data warehouse,
- Data Requirements and Data Description.

Goals for 2011: Temporal Database with All Available History

- Team: M. Böhlen, A. Bracher, F. Cafagna and A. Taliun.
- Duration: February - May.
- Key Tasks:
 - refinement of the design: categorical data and derived numerical attributes,
 - implementation of the database,
 - application to import the data from internal data formats,
 - import of all available data.
- Key Challenges:
 - new categorical data,
 - incomplete values for dates and locations.

Goals for 2011: On-line computation of up-to-date summaries

- Team: M. Böhlen, A. Bracher and A. Taliun.
- Duration: May - July.
- Key Tasks:
 - analysis and definition of the problem, i.e. what is a good summary and how to compute it?
 - development of the solution: start with a simple that works, but pursue the optimal;
 - integration into the feed database.

Goals for 2011: User interface will all current functionality

- Team: F. Cafagna and A. Taliun.
- Duration: August - December.
- Goal: migration to Model–View–Controller architecture that isolates application logic from the user interface and permits team development.
- Key Tasks:
 - re-engineering of the current code,
 - implementation,
 - testing,
 - documenting.

Goals for 2011: Data warehouse

- Team: M. Böhlen, A. Bracher and F. Cafagna.
- Duration: February - September.
- Goal: public data warehouse
- Key Tasks:
 - multidimensional model, stressing and completion,
 - implementation of the system on a server,
 - data import automatization and data analyses.

Goals for 2011: User interface to extract historical information

- Team: A. Bracher and A. Taliun.
- Duration: September - December.
- Key Tasks:
 - design 'good' queries that are required by the users,
 - graphical or textual presentation of the results.

Timetable for 2011

- February - March: implementation of the temporal database;
- March - May: data Import and refinement of the design;
- May - July: Computation of Up-To-Date summaries, complete data requirements;
- June-September: implementation of the data warehouse system on the server;
- August-December: implementation of the user interface:
 - first demo in September,
 - extraction of historical information in November,
 - final version in December.