ICA | INFORMATICA COMMUNICATIE ACADEMIE

Building an ER-model from verbalizations of facts Part 1: Sorting

HAN

Problems in ERM

Here: attempt to solve both problems using verbalizations of concrete examples of facts

Two main problems with classic ERM:

- Semantics not modeled
  - Semantics highly valued in practice
  - Comments on ETs and Atts often absent
  - Definitions: often only after the modeling
- **No good modeling method**
  - Most textbooks show WHAT to model
  - No textbook shows HOW to model

Why use verbalizations of facts?

Verbalizations of elementary facts:

- Capture the semantics of the data
  - Main issue in practice (>60% of design time)
- Offer a valuable alternative viewpoint
  - Indivisible units of information composed of several Atts that belong together (molecule)
  - Single Att: only a part of info (atom)
- Are independent of modeling technique
  - Do not change in model transformations: ERM, UML, Rel, …: same verbalizations
- Allow validation by the domain expert

How do they help a modeler?

Verbalizing concrete examples of facts:

- Makes the modeler understand the data
- Is done in dialogues between the modeler and the domein expert: no ‘ivory tower’ modeling
- Enables an arcane abstract ERM model to be built from familiar concrete facts
- Leads to a good and simple method to draw up an ERM diagram
- Enables easy validation of the model
- Enables adding semantics and examples to the diagram itself where convenient

Procedure to draw up an ERD

Steps 1 and 2 are not covered here.

Step 1: see the reader DM-RDS.

Step 2: see presentation Verbalizing.

Step 3: this presentation.

Step 4: see presentation Analyzing.

1. Collect concrete examples of facts
   - Use BPM as starting point
   - Make up examples if they don't exist (yet)
2. Verbalize these examples
   - With domain expert. Result: fact expressions.
   - Make the meaning as clear as possible
3. Sort expressions into Fact Types (FTs)
   - Same kind of expression: same FT
   - Order FTs with most components last
4. Analyze each FT (2 segments) and add the results to the ERD

Sorting fact expressions

Expressions of the same kind belong to a **Fact Type**.

Expressions of the same type have **components**: places where the text can vary.

FT4 has 2 components

FT6 has 3 components

```
FT4
Employee InsEd manages project P315.
    "    SmthE    "       "    P422.

FT6
Subproject 1 of project P315 starts on 20160201.
    "      2  "    "    P315    "    " 20160201.
    "      3  "    "    P315    "    " 20160208.
    "      1  "    "    P422    "    " 20160101.
```

Sorting fact expressions

Procedure for sorting:

- Place expressions of the same kind in the same Fact Type (FT)
  - 2 or 3 expressions per FT is enough
- Per FT: count the number of components
  - Component: place where text can vary
  - Highlight the components
- Order the FTs
  - FTs with the fewest components: first
  - FTs with the most components: last