Visual Specification of Multidimensional Queries based on a Semantic Data Model

Dr. Michael Böhnlein
Dipl.-Wirtsch.Inf. Markus Plaha
Dipl.-Inf. Achim Ulbrich-vom Ende

Computerbasiertes Entscheidungsunterstützungssystem für die bayerischen Hochschulen

T-Systems Nova
Entwicklungszentrum Darmstadt

Lehrstuhl für Wirtschaftsinformatik insb. Systementwicklung und Datenbankanwendung Otto-Friedrich-Universität Bamberg
Visual Specification of Multidimensional Queries
based on a Semantic Data Model
Scientific Community:
The specification of modeling approaches (1) and query languages (2) had great influence on our work.

(1) Modeling approaches:
- Dimensional Fact Model (DFM) from Golfarelli, Maio, Rizzi
- Multidimensional Unified Modeling Language (mUML) from OFFIS
- Fachkonzeptuelle Modellierung from Becker, Holten and Knackstedt
- Multidimensional Entity-Relationship-Model (M/ERM) from Sapia, Blaschka and Höfling
- Focus is on the documentation aspect.

(2) Query languages:
- They base on a formal data model using an algebra or a calculus.
- Emphasis is given to query optimization not to end-user friendliness.
- The relationship between modeling approaches and query languages is neglected.
Related Work

The industry has not yet adopted semantic data models from the scientific community.
Commercial products visualize business terms with elements of conventional GUIs.
Pragmatic approaches with the direct manipulation or drag and drop paradigm are applied.
Often a tree-view or a wizard-based approach to represent data is used.

Our Approach:

• Model:
  
  Modeling multidimensional schemas with a meaningful and implementation-independent semantic data model.

• Selection:
  
  Use of the semantic data model to select parts of a query.

• Query:
  
  Generating queries which lead to a grid or graph representation of the results.
Visual Specification of Multidimensional Queries based on a Semantic Data Model

Introduction | Related Work | SDWM | Query Specification | Tool Architecture | Summary

- ratio of enrolled students [faculty | university]
- ratio of cancelled study agreements (in %)
- ratio of cancelled study agreements [semester | previous semester]

- number of enrolled students
- number of cancelled study agreements

- field of study
- field of studies
- faculty
- Subject group
- university
- all

- semester
- all

- geographical origin
- abroad
- national
- district
- state
- country
- all

- reason for leaving
- university
- all

Query Specification Tool Architecture

Visual Specification of Multidimensional Queries based on a Semantic Data Model
Visual Specification of Multidimensional Queries based on a Semantic Data Model

Introduction | Related Work | **SDWM** | Query Specification | Tool Architecture | Summary

- **ratio of enrolled students** [faculty | university]
- **ratio of cancelled study agreements** (in %)
- **ratio of cancelled study agreements** [semester | previous semester]
- **number of enrolled students**
- **number of cancelled study agreements**

- **field of study**
  - faculty
  - Subject group
    - university
      - all

- **semester**
  - all

- **geographical origin**
  - abroad
    - national
      - district
        - state
          - country
            - all

- **field of studies**
- **time**
- **reason for leaving university**
  - all

**system of measurement view**
Visual Specification of Multidimensional Queries based on a Semantic Data Model

**Introduction**

**Related Work**

**SDWM**

**Query Specification**

**Tool Architecture**

**Summary**

---

**Dimensions**: field of study, faculty, Subject group, university, all

**Time**: semester, all

**Geographical Origin**: abroad, national, district, state, country, all

**Reason for Leaving University**: all

---

- Ratio of enrolled students [faculty | university]
- Ratio of cancelled study agreements (in %)
- Ratio of cancelled study agreements [semester | previous semester]
- Number of enrolled students
- Number of cancelled study agreements
- Ratio of cancelled study agreements [semester | previous semester]
- Number of enrolled students
- Number of cancelled study agreements
- Ratio of cancelled study agreements [semester | previous semester]

---

**SDWM**

System for visualized multidimensional queries based on a semantic data model.
Visual Specification of Multidimensional Queries based on a Semantic Data Model

Introduction  Related Work  SDWM  Query Specification  Tool Architecture  Summary

- ratio of enrolled students [faculty | university]
- ratio of cancelled study agreements (in %)
- ratio of cancelled study agreements [semester | previous semester]
- number of enrolled students
- number of cancelled study agreements

regular

dimension

- field of study
  - faculty
  - Subject group
    - university
      - all

- semester
  - all

- time

geographical origin

- abroad
  - national
    - district
      - state
        - country
          - all

reason for leaving university

reason for leaving

- all
Visual Specification of Multidimensional Queries based on a Semantic Data Model

Introduction  Related Work  SDWM  Query Specification  Tool Architecture  Summary

ratio of enrolled students  ratio of cancelled study agreements (in %)  ratio of cancelled study agreements [semester | previous semester]

number of enrolled students  number of cancelled study agreements

time  geographical origin

field of study

faculty  Subject group

university

all

semester  district

all  state

country

all

reason for leaving

reason for leaving university
Visual Specification of Multidimensional Queries based on a Semantic Data Model

- **Introduction**
- **Related Work**
- **SDWM**
- **Query Specification**
- **Tool Architecture**
- **Summary**

### Diagram

- **Ratio of enrolled students [faculty | university]**
- **Ratio of cancelled study agreements (in %)**
- **Ratio of cancelled study agreements [semester | previous semester]**

**Number of enrolled students**

**Number of cancelled study agreements**

**Unbalanced hierarchy**

**Field of study**
- **Faculty**
- **Subject group**
- **University**
- **All**

**Time**
- **Semester**
- **All**

**Geographical origin**
- **Abroad**
- **National**
  - **District**
  - **State**
  - **Country**
  - **All**

**Reason for leaving university**
- **All**

---

*Dr. M. Böhnlein, M. Plaha, A. Ulbrich vom Ende*

*13.11.2002*
Visual Specification of Multidimensional Queries based on a Semantic Data Model

Introduction

Related Work

SDWM

Query Specification

Tool Architecture

Summary

ratio of enrolled students [faculty | university]  

ratio of cancelled study agreements (in %)  

ratio of cancelled study agreements [semester | previous semester]

number of enrolled students

number of cancelled study agreements

field of study

faculty

Subject group

university

field of studies

semester

all

time

g encoded

distinct

g

g

national

district

state

country

all

geographical origin

abroad

all

reason for leaving

reason for leaving university

ratio of enrolled students [faculty | university]  

ratio of cancelled study agreements [semester | previous semester]

ratio of cancelled study agreements (in %)

number of enrolled students

number of cancelled study agreements

field of study

faculty

Subject group

university

field of studies

semester

all

time

reason for leaving

reason for leaving university

ratio of enrolled students [faculty | university]  

ratio of cancelled study agreements (in %)  

ratio of cancelled study agreements [semester | previous semester]
Visual Specification of Multidimensional Queries based on a Semantic Data Model

Introduction

Related Work

SDWM

Query Specification

Tool Architecture

Summary

Ratio of enrolled students [faculty | university]

Ratio of cancelled study agreements (in %)

Ratio of cancelled study agreements [semester | previous semester]

Number of enrolled students

Number of cancelled study agreements

Field of study:
- Faculty
- Subject group
- University
  - All

Time:
- Semester
  - All

Geographical origin:
- District
  - State
    - Country
      - All

Reason for leaving:
- All

Field of studies

Time

Geographical origin

Reason for leaving university

complex

measures

S: time

national

abroad

S:time

ratio of enrolled students

ratio of cancelled study agreements

ratio of cancelled study agreements
Visual Specification of Multidimensional Queries based on a Semantic Data Model

| Introduction | Related Work | SDWM | Query Specification | Tool Architecture | Summary |

- **Introduction**
- **Related Work**
- **SDWM**
- **Query Specification**
- **Tool Architecture**
- **Summary**

### Visual Specification of Multidimensional Queries

Based on a Semantic Data Model

**Complex Measures**

- **Ratio of enrolled students** ([faculty | university])
- **Ratio of cancelled study agreements (in %)**
- **Ratio of cancelled study agreements** ([semester | previous semester])

**Number of enrolled students**

- **Number of cancelled study agreements**

**Field of studies**

- **Field of study**
  - **Faculty**
  - **Subject group**
  - **University**
  - **All**

**Time**

- **Semester**
  - **All**

**Geographical origin**

- **Abroad**
- **National**
  - **District**
  - **State**
  - **Country**
  - **All**

**Reason for leaving university**

- **Reason for leaving**
  - **All**
Visual Specification of Multidimensional Queries based on a Semantic Data Model

Introduction

Related Work

SDWM

Query Specification

Tool Architecture

Summary

ratio of enrolled students [faculty | university]

ratio of cancelled study agreements (in %)

ratio of cancelled study agreements [semester | previous semester]

number of enrolled students

number of cancelled study agreements

field of study

faculty

Subject group

university

all

field of studies

semester

all

time

generational origin

abroad

national

district

state

country

all

reason for leaving university

reason for leaving university

all
Visual Specification of Multidimensional Queries based on a Semantic Data Model

Introduction

Related Work

SDWM

Query Specification

Tool Architecture

Summary

- ratio of enrolled students [faculty | university]
- ratio of cancelled study agreements (in %)
- ratio of cancelled study agreements [semester | previous semester]
- number of enrolled students
- number of cancelled study agreements

- field of study
- faculty
- Subject group
- university
  - all

- semester
  - all

- geographical origin
  - abroad
  - national
  - district
    - state
    - country
      - all

- reason for leaving university
  - all

- reason for leaving

- S:time

complex
measures
Visual Specification of Multidimensional Queries based on a Semantic Data Model

Introduction | Related Work | SDWM | Query Specification | Tool Architecture | Summary

- field of study
  - faculty
  - Subject group
  - university

- time
  - semester
  - all

- geographical origin
  - abroad
  - national
  - district
  - state
  - country
  - all

- reason for leaving
  - university

- ratio of enrolled students [faculty | university]
- ratio of cancelled study agreements (in %)
- ratio of cancelled study agreements [semester | previous semester]
- number of enrolled students
- number of cancelled study agreements
- additivity

S:time

S:time

S:time

visual specification of multidimensional queries based on a semantic data model
Visual Specification of Multidimensional Queries based on a Semantic Data Model

Introduction

Related Work

SDWM

Query Specification

Tool Architecture

Summary

multi-dimensional data structure 1 (hypercube)

ratio of enrolled students [faculty | university]

ratio of cancelled study agreements (in %)

ratio of cancelled study agreements [semester | previous semester]

number of enrolled students

number of cancelled study agreements

field of study

faculty

Subject group

university

all

field of studies

time

national

district

state

country

all

geographical origin

reason for leaving university

reason for leaving

all

S:time

S:time
Visual Specification of Multidimensional Queries based on a Semantic Data Model

Introduction

Related Work

SDWM

Query Specification

Tool Architecture

Summary

ratio of enrolled students [faculty | university]
ratio of cancelled study agreements (in %)

number of enrolled students

number of cancelled study agreements

ratio of cancelled study agreements [semester | previous semester]

field of study

faculty

Subject group

university

all

field of studies

all

time

semester

all

abroad

district

national

state

country

all

geographical origin

reason for leaving university

reason for leaving

[faculty | university]
Outline of problems for end-users

- **Wrong combination of levels and measures:**
  To avoid cross-joins, the end-users may not combine measures and dimension hierarchy levels from different cubes in one report.

- **Summarization of non-additive measures:**
  For non-additive measures either the corresponding hierarchy level has to be added to the report template or a single value of the level has to be added to the report filter.

- **Wrong combination of levels in parallel hierarchies:**
  Different levels of alternative paths in the same parallel hierarchy must not be added to the report.

- **Misinterpretation of report results with complex measures:**
  The selection of complex measures without selecting the corresponding hierarchy levels may result in misinterpretation.
Wrong combination of levels and measures (problem):

<table>
<thead>
<tr>
<th>semester</th>
<th>number of enrolled students</th>
</tr>
</thead>
<tbody>
<tr>
<td>winter term 2006/2007</td>
<td>73.096</td>
</tr>
</tbody>
</table>

Adding hierarchy level „*reason for leaving university*“

<table>
<thead>
<tr>
<th>semester</th>
<th>reason for leaving university</th>
<th>number of enrolled students</th>
</tr>
</thead>
<tbody>
<tr>
<td>winter term 2006/2007</td>
<td>cancelled after finally failed examination</td>
<td>73.096</td>
</tr>
<tr>
<td></td>
<td>cancelled after successful examination</td>
<td>73.096</td>
</tr>
<tr>
<td></td>
<td>cancelled for maternity</td>
<td>73.096</td>
</tr>
<tr>
<td></td>
<td>cancelled for university reasons</td>
<td>73.096</td>
</tr>
<tr>
<td></td>
<td>cancelled without examination</td>
<td>73.096</td>
</tr>
<tr>
<td></td>
<td>cancelled; examination not successful</td>
<td>73.096</td>
</tr>
<tr>
<td></td>
<td>changed university</td>
<td>73.096</td>
</tr>
<tr>
<td></td>
<td>conscripted for military services</td>
<td>73.096</td>
</tr>
<tr>
<td></td>
<td>interrupted or given up studying</td>
<td>73.096</td>
</tr>
<tr>
<td></td>
<td>not cancelled</td>
<td>73.096</td>
</tr>
<tr>
<td></td>
<td>other reasons</td>
<td>73.096</td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td>804.056</td>
</tr>
</tbody>
</table>
Wrong combination of levels and measures (solution):

„Reason for leaving“ is not selectable
Summarization of non-additive measures (problem):

<table>
<thead>
<tr>
<th>semester</th>
<th>number of enrolled students</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer term 2006</td>
<td>69.121</td>
</tr>
<tr>
<td>winter term 2006/2007</td>
<td>73.096</td>
</tr>
<tr>
<td>summer term 2007</td>
<td>69.141</td>
</tr>
<tr>
<td>winter term 2007/2008</td>
<td>71.220</td>
</tr>
<tr>
<td>summer term 2008</td>
<td>66.926</td>
</tr>
<tr>
<td>winter term 2008/2009</td>
<td>69.286</td>
</tr>
<tr>
<td>summer term 2009</td>
<td>59.539</td>
</tr>
<tr>
<td>winter term 2009/2010</td>
<td>63.006</td>
</tr>
<tr>
<td>summer term 2010</td>
<td>58.549</td>
</tr>
<tr>
<td>winter term 2010/2011</td>
<td>61.557</td>
</tr>
</tbody>
</table>

Removing hierarchy level „semester“

<table>
<thead>
<tr>
<th>number of enrolled students</th>
</tr>
</thead>
<tbody>
<tr>
<td>661.441</td>
</tr>
</tbody>
</table>
Summarization of non-additive measures (solution):

Automatic selection of "semester"
Wrong combination of levels in parallel hierarchies (problem):

University A:

<table>
<thead>
<tr>
<th>semester</th>
<th>subject group</th>
<th>number of enrolled students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>subject group of Economics</td>
<td>11361</td>
</tr>
<tr>
<td></td>
<td>subject group of Catholic Theology</td>
<td>951</td>
</tr>
<tr>
<td></td>
<td>subject group of Philology</td>
<td>9673</td>
</tr>
<tr>
<td></td>
<td>subject group of Mathematics and Natural Sciences</td>
<td>3661</td>
</tr>
</tbody>
</table>

winter term 2006/2007

University B:

<table>
<thead>
<tr>
<th>semester</th>
<th>faculty</th>
<th>number of enrolled students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Faculty of Philology</td>
<td>9567</td>
</tr>
<tr>
<td></td>
<td>Faculty of Economics</td>
<td>8456</td>
</tr>
<tr>
<td></td>
<td>Faculty of Mathematics and Natural Sciences</td>
<td>7468</td>
</tr>
<tr>
<td></td>
<td>Faculty of History and Geography</td>
<td>8036</td>
</tr>
</tbody>
</table>

winter term 2006/2007

Combining hierarchy levels "faculty" and "subject group" in one report:

<table>
<thead>
<tr>
<th>semester</th>
<th>subject group</th>
<th>number of enrolled students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Faculty of Philology</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Faculty of Economics</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Faculty of Mathematics and Natural Sciences</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Faculty of History and Geography</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Faculty of Social Science</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Faculty of Catholic Theology</td>
<td>0</td>
</tr>
</tbody>
</table>

winter term 2006/2007
Wrong combination of levels in parallel hierarchies (solution):

Either „faculty“ or „subject group“ is selectable.
### University B:

| semester        | faculty                                           | number of enrolled students | number of enrolled students [university] | ratio of enrolled students [faculty | university] |
|-----------------|---------------------------------------------------|----------------------------|------------------------------------------|-------------------------------------|
| winter term 2006/2007 | Faculty of Philology                             | 9567                       | 36548                                    | 26,18 %                             |
|                 | Faculty of Economics                              | 8456                       | 36548                                    | 23,14 %                             |
|                 | Faculty of Mathematics and Natural Sciences       | 7468                       | 36548                                    | 20,43 %                             |
|                 | Faculty of History and Geography                  | 8036                       | 36548                                    | 21,99 %                             |
|                 | Faculty of Social Science                         | 2089                       | 36548                                    | 5,71 %                              |
|                 | Faculty of Catholic Theology                      | 932                        | 36548                                    | 2,55 %                              |
| SUM             |                                                   | 36548                      | 36548                                    | 100,00 %                            |

#### Misinterpretation of report results with complex measures (problem):

Removing hierarchy level "faculty"
Misinterpretation of report results with complex measures (solution):

Automatic selection of „faculty“ and „university“
Visual Specification of Multidimensional Queries based on a Semantic Data Model

Introduction | Related Work | SDWM | Query Specification | Tool Architecture | Summary

logical architecture

physical architecture

implementation through program (PG)

server

client 1 ... client n

- data management D
- application logic A1
- presentation P

- JDBC DBMS
- Java class libraries
- RMI

- application logic A2
- Java class libraries
- SWING, JFC

- PG

- PG

- PG

visual specification of multidimensional queries based on a semantic data model
In this paper, we presented a new approach of a visual multidimensional querying tool.

**Goals of our work:**

- Helping end-users to transform their information demands into right syntax of a multidimensional query.
- Assistance in query specification by considering integrity constraints and therefore avoiding misformulated queries.
- First step in closing the loop between data modeling and query specification.
- Prototypical tool implementation ([http://ceus.uni-bamberg.de/jdwtoolsuite](http://ceus.uni-bamberg.de/jdwtoolsuite)).
Visual Specification of Multidimensional Queries based on a Semantic Data Model

Questions?

contact

michael.boehnlein@t-systems.com
markus.plaha@wiai.uni-bamberg.de
achim.ulbrich@wiai.uni-bamberg.de
Backup
## Benefits of SDWM:

- Mastering complexity
- Meta model
- Proven method
- Measurement system
- Additivity
- Support for the multidimensional paradigm
- Foundation of discussion for decision makers