Families of Database Schemas for Neuroscience Experiments

Larissa Cristina Moraes
Advisor: Kelly Rosa Braghetto

Institute of Mathematics and Statistics - University of São Paulo

May 6th, 2015

Agenda

Introduction
Neuroscience experiments

- Neuroscience experiments study the correlation between cerebral systems and normal or modified mental activity.
- Provenance data:
  - What are provenance? They are related to the experimental protocol used in the data collection and other orthogonal information (such as "when", "where' and "by who" data was collected).
  - Such metadata are essential for making judgements about data quality, integrity, and authenticity.
  - Provenance data are important for scientists when sharing and reusing experiment results.

Introduction
Experimental Data

Experimental Protocol:

- Components of protocol: task, stimulus, pause, instruction, questionnaire and block of components.

Data Collection:

- Types of data: Electroencephalography (EEG), Transcranial Magnetic Stimulation (TMS), Electromyography (EMG), Magnetic Resonance Imaging (MRI), Behavioral, etc.

Subject Types:

- Human and non Human subjects have specific characteristics.
- Human speaks language and has handedness.
- Non Human belongs to a specie, has a genus and a strain.

Statement of the Problem

Addressed problem:

- Digital representation and storage of neuroscience experimental data.

Challenges:

- Data variability makes it difficult to store data using traditional database models.
- There is a lack of patterns in scientific community for experimental data representation.
- The evolution of the structure of an "in-use" database is a very costly task which depends on IT specialists.

Proposed Solution
Objectives

- Create a family of conceptual database schemas for neuroscience experimental data.
- Create a notation that enables neuroscientists to easily extend the family, enriching it with the capacity of representing new experiment types.
- Develop a software tool to automate the creation or evolution of databases based on the family of schemas.

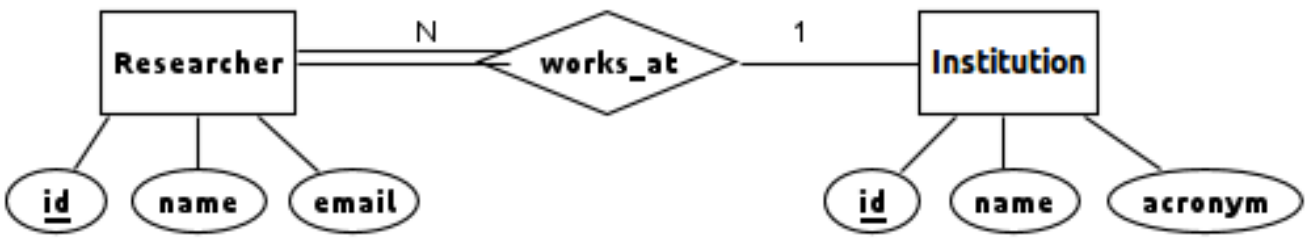
Proposed Solution
Other expected contributions

- Create customized conceptual database schemas, considering the specific needs of each research laboratory.
- Create patterns to represent neuroscience experimental data.
- Enable scientific community to share, reuse and reproduce data related to experiments.

Background
Conceptual schema

Modelling of conceptual schema of databases:

- Entity types: group of things.
- Attributes: characteristics of entities.
- Relationship type: associations between of entities.



Background
Software Product Line Engineering

SPL Definition

Software Product Line Paradigm promotes the reuse of software artefacts by managing the common and variable functionalities of a domain.

Phase 1: Domain Engineering

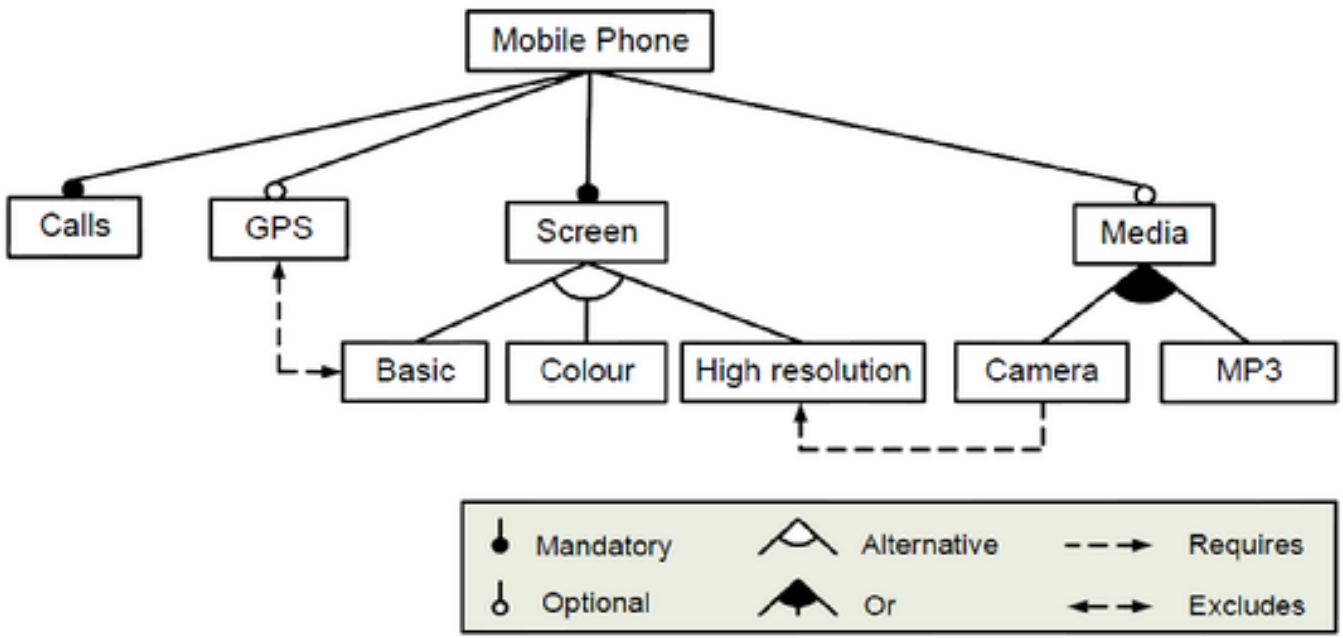
Common and variable functionalities are defined for the domain and an repository of reusable artefacts is built for the product line.

Phase 2: Application Engineering

The infrastructure created in previous phase is used as the basis to derive specific products of software.

Background
Software Product Line Example

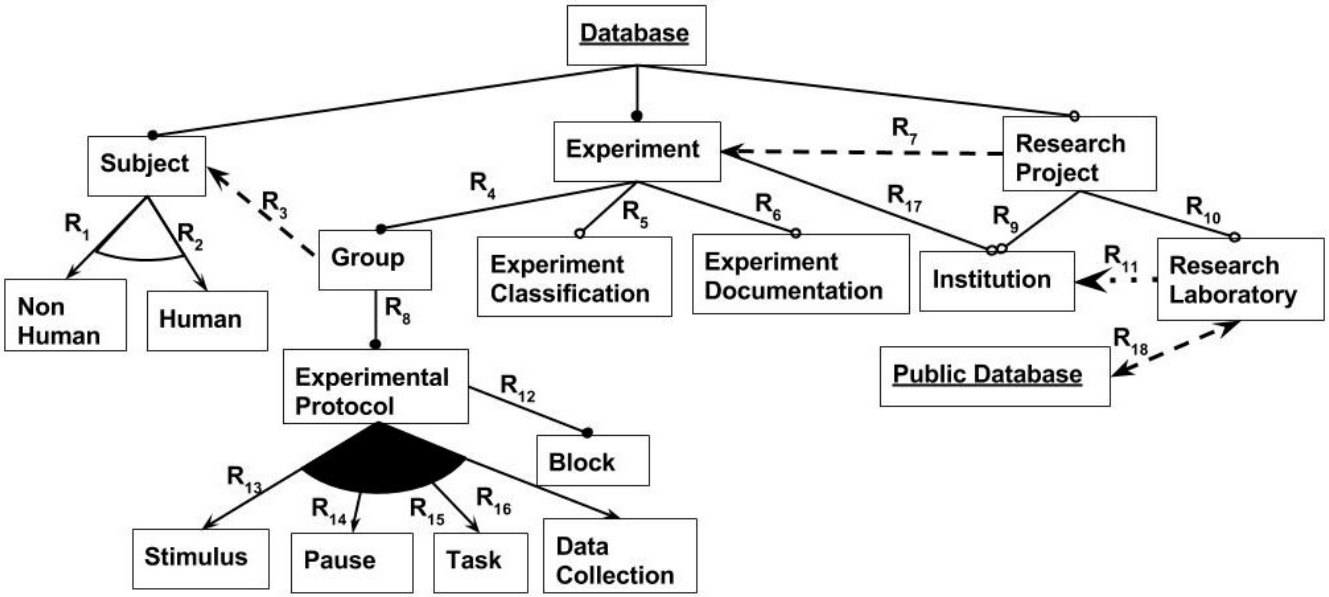
- Feature Diagram is commonly used to capture commonalities and variabilities between software applications.
- Example of mobile phone platform: Feature Diagram developed in Domain Engineering phase.



Methods and Approach
Main idea

- In this project, we defined an extension of the Feature Diagrams of SPL – the Database Feature Diagrams (DBFD) – specially to express data variability in database models.
- In our diagram, a module is a partition of a conceptual model and relations are used to express the dependencies and constraints existing among modules.
- Annotations were introduced to improve the expressive power of relations and represent which database modifications (like creating new relationship types and adding or removing attributes of existing entity types) should be made when a module is selected.

Methods and Approach
Example for Neuromat Project



R$_2$: ADD SPECIALIZATION subject_type BETWEEN (M Subject - E Subject) AND (M Human - E Human)
R$_8$: ADD RELATIONSHIP has_experimental_protocol BETWEEN (M Group - E Group) AND (M Experimental Protocol - E ComponentConfiguration) 1:1

Methods and Approach
Validation

- We will validate our proposal with groups of researchers of different laboratories associated to Neuromat.
- The quality of the proposed method will be evaluated by means of controlled experiments and case studies carried out in the context of NeuroMat project.
- Evaluation criteria: support to variability, flexibility, maintainability, and evolvability of the database models derived from the schema family created in the project.

The project in a nutshell

- **Objective:** Propose solutions to represent and store data of neuroscience experiments domain.
- **Challenges:**
  - Deal with variability of experimental data: experimental protocol, data collection and subject types.
  - Be "user-friendly" for neuroscientists.
- **Approach:** Adapt the Software Product Line paradigm to be used to represent data variability in terms of conceptual schemas of database.
- **Expected result:** Become a reference for representing neuroscience experimental data.

Thank you! Questions?

References


Bartholdt, J., Oberhauser, R., and Rytina, A. (2009)
Addressing data model variability and data integration within software product lines.


Buneman, P., Khanna, S., and Tan, W. C. (2001)
Why and where: A characterization of data provenance.


Khedri, N. and Khosravi, R. (2013)
Handling database schema variability in software product lines.


Kötter, R. (2001)
Neuroscience databases: tools for exploring brain structure–function relationships.


Van Der Linden, F. and Pohl, K. (2005)
Software Product Line Engineering: Foundations, Principles, and Techniques.