SPACEMAKER: PRACTICAL FORMAL SYNTHESIS OF TRADEOFF SPACES FOR OBJECT-RELATIONAL MAPPING

Hamid Bagheri, Kevin Sullivan, and Sang Son
University of Virginia Computer Science
How do expert engineers select a mapping from among the space of possible mappings?
Problem

- Developing object-relational (OR) mappings that achieve desired quality is difficult.

- Automatic generation of mappings using pure mapping strategies
  - Restricts the applicable mapping strategies
  - Leads to sub-optimal results

- Manual design of mixed mappings
  - Achieve much higher quality
  - Requires expertise in OR mapping strategies
Goals

Develop an approach and a tool capability that provide both the quality benefits of mixed mappings and the productivity benefits of automated synthesis.
Approach
A formal automated approach for exhaustive synthesis of mixed OR mappings and their classification into quality equivalence classes
Approach

- **Inputs:**
  - A formal object model
  - [optional] User-specified mapping constraints
- Generates the space of mappings
- Measures mappings in dimensions of six quality metrics
- Clusters mappings into quality equivalence classes
- Presents a candidate from each class to the engineer
**Spacemaker High-Level View**

**Formal object model**

```plaintext
module personObjectModel
open objectModel

one sig Person extends Class {}
attrSet = identifier+name
id=identifier
no parent
isAbstract = No

one sig name extends Attribute {}
type in string
```

**User-specified Constraints**

- open ORMStrategies
- open personObjectModel

- UnionSubclass[Manager]
- genericMixedStrategy[Clerk, Student, Employee]

**Candidates from each class**

- Object-Relational Mapping Strategies (formalized)

- Measuring & Clustering

**Constraint Solver**

We formalize OR mapping strategies as Alloy predicates

- ORM strategies to manage association relationships
- ORM strategies to manage inheritance relationships
- Fine-grained strategies applicable to classes within an inheritance hierarchy

Alloy is a first-order specification language based on the set theory that has been optimized for automated analysis.
UnionSubclass Strategy
As the Alloy predicate

\[
\text{pred UnionSubclass}[c :: \text{Class}]{
  c \text{ in (isAbstract.No)} \Rightarrow \{ \\
  \text{one Table <:c.^tAssociate} \\
  \}
}
\]

If the given class is not abstract, it will be assigned to a separate table.
UnionSubclass Strategy
As the Alloy predicate

The table includes relational fields corresponding to attributes of the associated class and its inherited attributes

\[
\begin{array}{l}
(c.\text{isAbstract}=\text{No}) \Rightarrow \{ \\
\quad \text{all } a: \text{Attribute} \mid a \text{ in c.attrSet} \Rightarrow \{ \\
\quad\quad \text{one } f: \text{Field} \mid f.fAssociate=a \\
\quad\quad \&\& f \text{ in } (c.\sim t\text{Associate}.\text{fields}) \} \\
\}
\end{array}
\]

\[
\begin{array}{l}
(c.\text{isAbstract}=\text{No})\&\&(c.^\text{parent} \neq \text{none}) \Rightarrow \{ \\
\quad \text{all } a: \text{Attribute} \mid a \text{ in c.^\text{parent}.attrSet} \Rightarrow \{ \\
\quad\quad \text{one } f: \text{Field} \mid f.fAssociate=a \&\& \\
\quad\quad f \text{ in } (c.\sim t\text{Associate}.\text{fields}) \} \\
\}
\end{array}
\]

To retrieve an individual object, only one table needs to be accessed
UnionSubclass Strategy
As the Alloy predicate

```
pred UnionSubclass[c : Class]{
  c in (isAbstract = No) => {
    all a : Attribute | a in c.attrSet => {
      one f : Field | f.fAssociate = a
        && f in (c.^tAssociate.fields)
    }
  }
  (c.isAbstract = No) && (c.^parent != none) => {
    all a : Attribute | a in c.^parent.attrSet => {
      one f : Field | f.fAssociate = a
        && f in (c.^tAssociate.fields)
    }
  }
  (c.^tAssociate).foreignKey = none
...}
```
Measuring Impacts of OR mappings

- Measure each synthesized alternative in dimensions of six quality metrics

- Using a set of metrics suggested by Holder et al. [Model-Based Software and Data Integration, 2008] and Baroni et al. [International Conference on Enterprise Information System, 2005]

- Develop a set of queries to execute over synthesized alternatives
Number of Corresponding Relational Fields (NCRF)

- NCRF metric: the extent of change propagation for a given OR mapping

- Definition: given a class C, NCRF(C) return number of relational fields in all tables that correspond to each non-inherited, non-key attribute of C

- Alloy Query to measure the NCRF metric:

\[
NCRF(C) = \#(\text{C.attrSet} - \text{C.id}).\sim f\text{Associate}.\sim \text{fields}
\]
Clustering

- Classify space of mappings into quality equivalence classes
- Each equivalence class: mapping alternatives that exhibit the same characteristics w.r.t. quality metrics
  - Alternatives $a_1$ and $a_2$ are equivalent:
    \[
    \text{value}(a_1, m_i) = \text{value}(a_2, m_i), \text{ for all metrics } (m_i)
    \]
- Select a single candidate from each equivalence class for presenting to end-users
Model Splitting

- To address scalability issue, split object model into submodels
- Key idea: For association with cardinality of many-to-many, there is just one applicable mapping strategy
- Split object model by removing all bridges of type many-to-many association
There are two such bridges:

\(< Product; Asset >\) and \(< Item; Product >\)

By removing bridges we obtain three smaller sub-graphs

The gain comes from the reduction in sizes of constraint solving problems
Evaluation
Hypotheses

- **Feasibility:**
  - It is feasible to formalize the correctness constraints for object-relational mapping strategies
  - To automate the formal synthesis of an exhaustive set of mapping alternatives

- **Performance:** performance of technology implementation based on a bounded model checker is reasonable
Two applications of mixed mapping strategies over a common object model

- White Boxes $\rightarrow$ Classes
- Grey-title boxes $\rightarrow$ Mapped Tables
- Highlighted labels $\rightarrow$ Applied Mapping strategies
Spacemaker generates the mapping diagram
According to user-specified constraints

module personObjectModel
open objectModel

one sig Person extends Class {}
attrSet = identifier+name
id=identifier
no parent
isAbstract = No

one sig name extends Attribute {}
type in string

Person object model
in Alloy

open ORMStrategies
open personObjectModel

UnionSubclass [Manager]
JoinedSubclass [Clerk]
UnionSuperclass [Employee]
UnionSubclass [Student]

User-specified constraints

Automatically generated
mapping diagram according to
User-specified constraints
Table 1 is associated to Person and Employee classes with respect to the UnionSuperclass strategy. Separate tables are assigned to Student and Manager classes, according to the UnionSubclass strategy. Table 3 is assigned to the Clerk class because of the JoinedSubclass strategy.
Spacemaker generates a new mapping diagram by changing the user-specified constraints

```
module personObjectModel
open objectModel

one sig Person extends Class {
  attrSet = identifier+name
  id = identifier
  no parent
  isAbstract = No
}

one sig name extends Attribute {
  type in string
}
```

**Person object model in Alloy**

```
open ORMStrategies
open personObjectModel

JoinedSubclass[Manager]
UnionSuperclass[Clerk]
UnionSuperclass[Student]
```

**User-specified constraints**

**Automatically generated Mapping diagram**
Spacemaker generates the whole space of mappings

```alloy
module personObjectModel
open objectModel

one sig Person extends Class {}
  atStr = identifier+name
  id=identifier
  no parent
  isAbstract = No

one sig name extends Attribute {}
  type in string

open ORMStrategies
open personObjectModel

genericMixedStrategy[Manager,Clerk,Student,Employee]

genericMixedStrategy
  lets model finder
  choose for each class
  any of the fine-grained strategies

Exhaustively generated space of mapping alternatives along with Quality Measures
```

Person object model in Alloy
Multi-dimensional quality measures

- If the designer prefers the resource utilization, the mapping (a) would be a better option.
- The ANV metric represents additional storage space in terms of null values.
- With respect to the ANV metric, the mapping (b) requires more wasted space.
Multi-dimensional quality measures

- If the designer prefers maintainability and performance, the mapping (b) would be a better option.
  - The NCRF metric reflects the effort required to adapt the relational schema.
  - The TATI metric is a performance indicator of database queries.
  - Mapping (a) negatively affects the NCRF and TATI metrics.
Without decomposition, the Alloy analyzer ran out of memory before generating the whole tradeoff space
Case study 2
An ecommerce object model

- Decompose the object model into three submodels
- Tradeoff space consists of hundreds of thousands of mappings
- Spacemaker generates mapping space < an hour
- Certainly a reasonable performance

<table>
<thead>
<tr>
<th>ObjectModel</th>
<th>Solutions</th>
<th>Eq.Classes</th>
<th>$T[\text{min}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product</td>
<td>137,000</td>
<td>67</td>
<td>46</td>
</tr>
<tr>
<td>Asset</td>
<td>124,000</td>
<td>31</td>
<td>$&lt; 1$</td>
</tr>
<tr>
<td>Order</td>
<td>93,000</td>
<td>31</td>
<td>16</td>
</tr>
</tbody>
</table>

e-commerce experiment performance statistics
Thank You