Model Comparison
A Systematic Mapping Study

Lucian Gonçales, Kleinner Farias, Murillo Scholl, Toacy Oliveira, Maurício Veronez

PIPCA – Universidade do Vale do Rio dos Sinos (UNISINOS)
lucianjosegoncales@gmail.com

SEKE 2015, 6-8 July 2015, Pennsylvania-Pittsburgh, USA
Model comparison is the activity of comparing at least two input models.

It can be used for Matching, calculate similarity, clone detection, pattern detection...

Then, this activity plays a pivotal role in Model Driven Engineering (MDE):
Introduction

Main purpose: Development oriented to a specific domain
Current approaches still is not providing a precise and large-scale computation in synchronizing and matching models

Then....
Introduction

Craftsmanship era!
Introduction

• A comprehensive understanding about the state-of-the-art is crucial for evolving the current comparison techniques;

• A systematic mapping study to
  • (1) scrutinize those contributions produced over time,
  • (2) characterize previously published model comparison approaches
Study Methodology

• **Search strategy for comparison approaches**
  - Definition of terms to form Search Strings for performing searches in the main digital libraries

• **Inclusion and exclusion criteria**
  - Search was limited to studies published in electronic digital libraries;
  - No restriction on the publication year of studies until November 2014;
  - Papers and studies which not focus on model comparison;
  - Duplicated studies returned by different search engines; and
  - Papers and works that focus in low-level comparison (XML, source code and text).

• **Classify extracted data**
  - (1) publication date, publication fora, and search engine; and
  - (2) basic attributes of studies: main author and title; and finally
  - (3) information related to research questions
RQ1: What are the types of diagrams addressed by comparison techniques?

• Find out the types of diagrams that comparison techniques support;
• Reveal the diagrams that have been considered important
RQ1 - Results

Inside MDE, capability to dealing with many kinds of model are required

Majority

The most frequent UML Diagram
Research Question 2

• **RQ2:** What are the data structures commonly used in the comparison algorithms?

  • Pinpoint which data structures are used in the comparison algorithms
RQ2 - Results

- Graph: 58%
- Tree: 15%
- Inferency: 3%
- Other: 25%
• What are the types or categories used for evaluating diagrams in similarity approaches?
  • Understand the different aspects is required to evaluate diagrams:
    • (1) Structural: compare diagrams considering modules and its relationships;
    • (2) Syntatic: comparing taking account the sintaxes of diagrams;
    • (3) Semantic: comparing diagrams considering the meaning;
    • (4) Layout: the comparison approaches aim at view issues;
    • (5) Lexical: implement a name-based model comparison;
    • (6) Multi-Strategy: approahes combine at least two comparison strategies to improve comparison results.
RQ3 - Results

The majority of papers focuses on comparing structural aspects, with:

- Structure: 21 papers
- Syntactic: 1 paper
- Semantic: 3 papers
- Lexical: 3 papers
- Layout: 1 paper
- Multi-Strategy: 11 papers

Low evolving on these aspects.
• How Fine-Grained are the comparison techniques?
  • Grasp how accurate and detailed are the comparison techniques in relation to model signatures:
    • Coarse-grained: low level of detail
    • Partial: a consensus
    • Fine-grained: high level of detail
RQ4 - Results

![Bar Chart]

- **Partial**: 29
- **Coarse-grained**: 8
- **Fine-grained**: 3
Research Question 5

• What are the comparison types?
  • Explore what kind of comparison the techniques are responsible for:
    • (1) Matching: Find the correspondent element in another diagrams
    • (2) Similarity: the score of correspondence between elements or between the whole diagram.
RQ5 - Results

![Bar chart showing the results for Matching and Similarity. The chart indicates a high similarity score of 28 compared to a lower matching score of 12.]
Which empirical strategies are used to evaluate the comparison techniques?

- Check the empirical strategies used to evaluate comparison techniques
  - (1) Evaluation research;
  - (2) Proposal of solution;
  - (3) Philosophical paper;
  - (4) Personal Experience;
  - (5) Opinion paper.
RQ6 - Results

- **Proposal of Solution**: 31
- **Evaluation Research**: 4
- **Philosophical Paper**: 4
- **Personal Experience**: 1
Research Question 7

• Is the approach automatic, semi-automatic or manual?
  • To Summarize the autonomous level of approaches.
RQ7 - Results

![Bar chart showing results for Automatic, Semi-automatic, and Manual categories with counts of 27, 12, and 1 respectively.](image-url)
## Quantity of papers per Event/Journal

<table>
<thead>
<tr>
<th>Publication Place</th>
<th>Quantity of approaches</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE/ACM International Conference on Automated Software Engineering (ASE)</td>
<td>4</td>
<td>10%</td>
</tr>
<tr>
<td>IEEE Transactions on software Engineering</td>
<td>3</td>
<td>8%</td>
</tr>
<tr>
<td>European Software Engineering Conference and the ACM SIGSOFT</td>
<td>2</td>
<td>5%</td>
</tr>
<tr>
<td>International Conference on Software Maintenance (ICSM)</td>
<td>2</td>
<td>5%</td>
</tr>
</tbody>
</table>
Publications grow in a time interval of Three years
## Rank of authors publications

<table>
<thead>
<tr>
<th>Author</th>
<th>Quantity of Papers</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhenchang Xing</td>
<td>3</td>
<td>8%</td>
</tr>
<tr>
<td>Christian Gerth</td>
<td>2</td>
<td>5%</td>
</tr>
<tr>
<td>Hamza Onoruoiza Salami</td>
<td>2</td>
<td>5%</td>
</tr>
<tr>
<td>Kleinner Farias</td>
<td>2</td>
<td>5%</td>
</tr>
<tr>
<td>Mark van den Brand</td>
<td>2</td>
<td>5%</td>
</tr>
<tr>
<td>Segla Kpodjedo</td>
<td>2</td>
<td>5%</td>
</tr>
<tr>
<td>Shiva Nejati</td>
<td>2</td>
<td>5%</td>
</tr>
</tbody>
</table>
Combined research questions

Year

Research method

Diagrams’ type
Threats to validity

(1) difficulty to relate all works to the topic due the constant changes in publications;

(2) the conduction of data extraction of the papers, such as:

(1) The search string we used has the main terms such as “model” and “matching”. However, “matching” and its synonyms (comparison, similarity, etc.) are generic and this string retrieved broad results;

(2) The inclusion of thesis and dissertations published online that are not peer reviewed and,

(3) The limitation to the main six search engines defined in the SMS planning.
Conclusion

• This paper identified and classified publication fora, and performed thematic analysis of the existing literature in model comparison.

• The most studies have concentrated more effort on producing generic comparison techniques:
  • 1º - There is not a widely-adopted modeling language in industry.
  • 2º - The wide variations of modelling notations and diagrams types, it would be challenging to provide an approach that can have a broad adoption.
  • 3º - Model comparison is not a trivial task to deal with.
References


