Semi-Automatic Identification of Features in Requirement Specifications

Ekaterina Boutkova, Dr. Frank Houdek
Overview

- Requirement Documents at Daimler Passenger Car Development (PCD)
- Identification of Features in the existing Specifications
- Evaluation and Lessons Learned
Requirement Documents Hierarchy at Daimler PCD

Products

Vehicle

System

Component

Requirement Documents

Semi-Automatic Identification of Features in Requirement Specifications, RE’2011 / 2.9.2011
Properties of Specifications at Daimler PCD

*Specification Style*

- Usually no formal specification
- Mainly natural language (German, English)
- Where appropriate: Tables, figures, etc.
- Specification template driven document
- Documentation on the different levels of abstraction
- Typically many requirements changes during a project
- Documentation of many variants
- Use of IBM Ration DOORS
- Tagging of requirements: Requirement, Information, Heading

*Specification Size*

<table>
<thead>
<tr>
<th>Number of pages</th>
<th>60</th>
<th>600</th>
<th>2,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of DOORS objects</td>
<td>1,000</td>
<td>50,000</td>
<td></td>
</tr>
<tr>
<td>Number of referenced documents</td>
<td>30</td>
<td>300</td>
<td></td>
</tr>
</tbody>
</table>

*Template Size*

| 1.500 | 150 S. |
Reuse of Requirements

Reuse approaches
- Copy and adapt
- Extension of existing specification
- Requirements pool

Documentation of variability
- In the requirements text
- Matrix approach (simple)
- Matrix approach (extended)

Shortcomings
- High manual effort
- High probability of error
- No means to check the correctness of mapping of requirements to products
### Feature-Based Variability Management

#### Feature Model

<table>
<thead>
<tr>
<th>Feature</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sun Visor</strong></td>
<td></td>
</tr>
<tr>
<td>Décor</td>
<td>The décor material of the sun visor shall be leather.</td>
</tr>
<tr>
<td>PVC</td>
<td></td>
</tr>
<tr>
<td>Drapery</td>
<td>The component shall comply all requirements of the design department</td>
</tr>
<tr>
<td>Leather</td>
<td>The light shall be activated through the turn down of sun visor.</td>
</tr>
</tbody>
</table>

#### Feature Specifications

- **Leather**
- **Light**

#### Key Points

- **Common requirements are not mapped**
- **Variable requirements are mapped to relevant features**
- **A product definition (i.e. product variant) forms a filter for the requirements specification**
Benefits of using Feature-Based Variability Management: Example Component Sun Visor

Basis: 400 requirements, 3 products (variants)

Matrix approach:

- Effort for a new product (variant): 400 assignments

Feature-based variability management:

- Effort for a new product (variant): 11 assignments

* Nonrecurring effort
Challenge: Identification of Features

- **Manual Identification (by comparing specifications)**

  *Specification A:*
  “200ms after DOORS_LOCK_SIGNAL has been detected, the signals IND_FRONT and IND_REAR have to be sent.”

  *Specification B:*
  “200ms after DOORS_LOCK_SIGNAL has been detected, the signals IND_FRONT, IND_REAR, and **IND_MIRROR** have to be sent.”

  - What is different? **IND_MIRROR**
  - Why? Vehicle has additionally an indicator light in the exterior mirror

- **Semi-automatically Identification** ➔ **Next slides**
Feature Identification Approach MIA

Semi-automatic approach for feature identification in the existing specifications

Uses of methods of the lexical analysis

PREPARATION

- Selection of a specification document
- Definition of criteria for analysis:
  - All Objects
  - Requirements only
  - Variable requirements only

MIA-ALGORITHM

1. Feature Candidates
2.
3.
4.

EVALUATION OF OUTPUT-DATA

- Selection of features
- Extension of stop-word list
- Extension of dictionary

BUILDING FEATURE MODEL

- Building a feature model
- Identification of relations between individual features
Feature Identification Approach MIA: MIA Algorithm

1. **TAGGING OF NOUNS**
   - Uses Stanford Part-of-Tree-Tagger for identification of nouns

2. **LEMMATIZATION**
   - Dictionary with 18,789 nouns and proper nouns in the basic form
   - Table with nouns ending

3. **DUPLICATE ELIMINATION**

4. **CLEANUP OF STOP WORDS**
   - Basis stop word list with 2,701 nouns and acronyms build from system and component specification templates
### Application of MIA: First Results

Does MIA find all relevant features? How many false positives are found?

<table>
<thead>
<tr>
<th>Number of</th>
<th>Manual analysis</th>
<th>All Objects</th>
<th>Requirements only</th>
<th>Variable req. only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>R</td>
<td>P</td>
<td>#</td>
</tr>
<tr>
<td><strong>Data set 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objects</td>
<td>727</td>
<td>-</td>
<td>-</td>
<td>543</td>
</tr>
<tr>
<td>Candidates</td>
<td>414</td>
<td>-</td>
<td>-</td>
<td>365</td>
</tr>
<tr>
<td>Features</td>
<td>9</td>
<td>8</td>
<td>0,89 0,02</td>
<td>8</td>
</tr>
<tr>
<td><strong>Data set 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objects</td>
<td>1397</td>
<td>-</td>
<td>-</td>
<td>1124</td>
</tr>
<tr>
<td>Candidates</td>
<td>287</td>
<td>-</td>
<td>-</td>
<td>213</td>
</tr>
<tr>
<td>Features</td>
<td>104</td>
<td>56</td>
<td>0,56 0,02</td>
<td>56</td>
</tr>
</tbody>
</table>

Recall (R) = \( \frac{\text{Feature candidates } \cap \text{ Feature}}{\text{Features}} \)

Precision (P) = \( \frac{\text{Feature candidates } \cap \text{ Feature}}{\text{Feature candidates}} \)
Application of MIA: First Results

How many nouns are identified by the tagger?

In a sample of 100 randomly selected requirements, in 84 all nouns had been identified correctly; in the remaining 16 requirements most nouns had been identified correctly. Main Problem: Bi-lingual requirements

Which features are hard/impossible to find?

Numbers (e.g. part numbers, catalog numbers)
Implicit features, i.e. features that are not mentioned explicitly

What are the false positives?

Often: Common features (that are not needed in our approach)

What are potentials for improvement?

Considering synonyms
Clustering of nouns with identical word stem
Enhancement of dictionary
Experiences from MIA application in practice

- Requirements experts at Daimler use MIA in building feature models.
- In a typical specification, the candidate list contains about 150 entries (variable requirements only); here, identifying stop words and similar words (identical word stem) does not disturb significantly.
- Evolution of word lists from release of MIA until today:

<table>
<thead>
<tr>
<th>Dictionary</th>
<th>Stop word list</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.789</td>
<td>2.701</td>
</tr>
<tr>
<td>26.523</td>
<td>4.212</td>
</tr>
</tbody>
</table>

- Hard challenge: Identification of reasons behind variation = Feature.
- Candidate lists are considered as a valuable input for discussions with specification authors.
Empirical Evaluation of MIA Effort Benefit

Approach:
- Data sets
- Manual identification of features
- Comparison of feature identification effort
- Identification of feature candidate list; manual review of list

Results:

<table>
<thead>
<tr>
<th>Time effort</th>
<th>Manual only</th>
<th>With MIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set 1</td>
<td>1:00</td>
<td>00:50</td>
</tr>
<tr>
<td>Data Set 2</td>
<td>4:00</td>
<td>3:40</td>
</tr>
</tbody>
</table>

Differences are not statistically significant, but the trend in the student experiment correlates with subjective evaluation in practice.