Supporting the Development of Cyber-Physical Systems with Natural Language Processing: A Report

Andreas Vogelsang, Kerstin Hartig, Florian Pudlitz, Aaron Schlutter, Jonas Winkler

Technical University of Berlin
18.03.2019
Automated Systems Engineering Technologies

Our research focus: We research and develop technologies to support system engineers and automate time-consuming or error-prone tasks and process steps.

AI for RE

Model-based Systems Engineering

Validation by Simulation

Lead: Andreas Vogelsang
PhD students:
- Tiago Amorim
- Florian Brokhausen
- Patrick Ebel
- Kerstin Hartig
- Florian Pudlitz
- Stefan Rulewitz
- Aaron Schlutter
- Jonas Winkler
- Florian Wiesweg

Website: aset.tu-berlin.de

+ 3 student assistants
Cyber-Physical Systems (CPS)

- Observe environment by sensors, influence by actuators
- Composed of mechanics, electronics, software
- Software most important and most critical

- Development is interdisciplinary
  - application domains
  - engineering disciplines

- Majority of development information is expressed in natural language

- Development is driven by strong safety and security constraints

Natural Language in Cyber-Physical Systems

Information spread over hundreds of documents with thousands of entries (e.g., specification repository of a telematics system of a modern vehicle: ~30,000 documents, ~2.5 million textual entries)
Past and Current Research on NLP for CPS Development
Automatic Knowledge Extraction

Glossary Term Extraction

- Requirement Statement as user story
- Tokenization
- POS-Tagging
- Chunking
- Lemmatization
- Relevance Filtering
- Specificity Filtering

Linguistic Processing using NLP Pipeline

Identification of Glossary Term Candidates

Statistical Filtering

Reduction of Glossary Term Candidates

Knowledge Graph Extraction

- Any textual requirements
- Filtering
- Tokenization
- POS-Tagging
- Lemmatization
- Semantic Role Labeling
- Graph Building

Gemkow, Conzelmann, Hartig, Vogelsang: “Automatic glossary term extraction from large-scale requirements specifications”, RE’18
Automatic Knowledge Extraction

Schlutter, Vogelsang: “Knowledge Representation of Requirements Documents Using Natural Language Processing”, NLP4RE’17

77 documents with 45,092 objects
73,878 nodes (i.e., concepts used)
134,866 edges (i.e., concept relations)
Expert Systems for Hazard and Risk Analysis

Safety Expert

Structured Query

Give me all elements of type <X> related to element <Y>

HaRa repository

Representation

Graph-based query

Semantic knowledge graph

Retrieval

Spreading Activation

Result

Filter elements of type <X>

Recommendations

<table>
<thead>
<tr>
<th>Elements</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

155 HaRa documents:
600 functions; 1,700 malfunctions; 4,200 hazards; 540 safety goals

Expected Precision (EP): 0.66
Automatic Requirements Classification

The duration until the switch is recognized as hanging must be a configurable parameter.

The component conditionally drives an external fan. This fan is required for active ventilation of the headlight.

Trained on datasets with:

- 36,104 requirements (English)
- 68,178 requirements (German)
- Requirement/Information: 60:40

~85% accuracy

+ Context

~90% accuracy

Winkler, Vogelsang: “Automatic Classification of Requirements Based on Convolutional Neural Networks”, AIRE’16
Future Research on NLP for CPS Development
Connecting NL Requirements and Simulation

1. **Automatic windscreen wiping** is activated in case of **detected rainfall**
2. The motor stops in the following cases
   - Crash 1%
   - Start/Stop 70%
   - Manual operation 29%
   - Key out of vehicle 0%
3. **Procedure of function**

   - **Cond. 1**: Check
   - **Cond. 2**: Start
   - **Cond. 3**: Stop

4. **Scenarios**
   - Object detected 12x
   - Vehicle stands 38x
   - Ø 3 sec.

**Log file**
1. Ign_on
2. v = 50 km/h
3. Oncoming traffic
4. Rain starts
5. WS_Whip sent
6. Signal X
7. Break – Obstacle detected

Pudlitz, Vogelsang, Brokhausen: “A Lightweight Multilevel Markup Language for Connecting Software Requirements and Simulations”, REFSQ’19
The component conditionally drives an external fan. This fan is required for active ventilation of the headlight.

The duration until the switch is recognized as hanging must be a configurable parameter.

Natural Language in CPS

Past research

Future research

We **research and develop** technologies to **support** system engineers and **automate** time-consuming or error-prone tasks and process steps.

@andivogelsang

andreas.vogelsang@tu-berlin.de