Natural Language Processing with Process Models

Jan Mendling (Vienna University of Economics and Business), Henrik Leopold (Kühne Logistics University), Lucineia Heloisa Thom (Federal University of Rio Grande do Sul), Han van der Aa (Humboldt-Universität zu Berlin)
BPM Lifecycle

1. Process identification
2. Process architecture
3. Process discovery
- Conformance and performance insights
- As-is process model
4. Process analysis
- Insights on weaknesses and their impact
5. Process monitoring and controlling
- Executable process model
6. Process implementation
- To-be process model
7. Process redesign
Current versus desirable quality

- Sub processes with inconsistent connection to main process
- Message flow arcs used on incorrect nodes
- Models with multiple merges
- Models with deadlocks
- Models with incorrect syntax
- Models with excessive diagram size
- Models with overlay of edges and nodes
- Models with inconsistent incoming and outgoing behavior
- Models with incorrect modeling direction
- Models with inappropriate spacing
- Data objects with no link to glossary
- Roles with no link to glossary
- Gateways with non-compliant labeling
- Activities with non-compliant labeling
- Events with non-compliant labeling

585 BPMN 2.0 process model from six companies

Leopold et al. 2015
25 Challenges of Semantic Process Modeling

Mendling et al. 2015
Challenges related to labels

C1: Identify Label Grammar
   Read Label → Read Label

C2: Refactor Label Grammar
   Label Reading → Read Label

C3: Disambiguate Label Terms
   Call Bank → Call Bank

C4: Refactor Label Terms
   Call Bank → Contact Financial Institution

C5: Auto-Complete Label
   Bank → Contact Bank

C6: Calculate Label Specificity
   Call Bank → Specificity

C7: Calculate Label Similarity
   Call Bank → Contact Bank → Similarity

Mendling et al. 2015
C1: Identify Label Grammar

Plan Data Transfer to EC-PCA

Plan Integration of Profit Centers

Action = Plan?

Manual Profit Center Planning

Action = Plan
Challenges related to models

C8: Discover Label Mapping
I read the label

C9: Identify Semantic Fragment

C10: Identify Fragment Name

C11: Unfold Label to Structure
Do A, B, C

C12: Transform Model to Text
A -> B -> C

C13: Transform Text to Model
First do A, then do B, then do C

C14: Verify Model Correctness
A -> B -> C

C15: Validate Model Completeness
A -> B -> C

C16: Auto-Complete Model
A -> B -> C

C17: Calculate Model Specificity
Specificity

C18: Translate Model
A -> B -> C

C19: Calculate Model-Text Consistency
Consistency

Mendling et al. 2015
Main challenges

- Syntactic Leeway
  - Changing active and passive voice of input text
- Atomicity
  - Activities can be split across sentences
- Relevance
  - Relative clauses, example sentences or meta-statements should not be translated to model elements
- Referencing
  - Anaphora, textual links

Evaluation

- Test set of 47 text-model pairs
- Average translation accuracy is of 77%

Friedrich et al. 2011
Translation from Process Model to Text

- Main Challenges
  - Text Planning
    - text structuring
  - Sentence Planning
    - lexicalization and message refinement
  - Surface Realization
    - interfacing with established realizers
  - Flexibility
    - addresses variations of input data and adaptation of output

Leopold, 2015
Conformance Checking between Process Models and Text

Main challenge
- Check the conformance of process models and text
- Align text labels of process models

Achievements
- Recorded process executions is compared with natural language specifications of processes
- Ambiguity detected with probabilistic conformance checking
• Process – oriented text
  • Structured
  • Capable of maintaining the maximum information related to the business process
  • Is in conformance with BPMN 2.0
Process-Oriented Text Generation from Natural Language Text

Silva, 2018
Tool Overview

The process starts when the technician performs an evaluation. Next, one of the 3 alternative procedures is performed. If "a software problem", the technician formats the computer. If "a hardware problem", the technician makes the no modification to the computer. If "founds no problem", the technician replaces the part and then fills out the part replacement form. In any case, the technician completes the repair form. Finally, the process ends.
How a Process –Oriented Text must be Structured

- Describe the text as a sequence of facts
- Use no more than 5 sentences
- Use passive voice
- Make explicit in the text splits and joins
- Describe all the paths from the beginning of the process until a gateway. After, describe the next paths

Silva, 2018 based on Dumas, 2013
Process – Oriented Text Generation from Natural Language Text

- Main characteristics of the approach
  - Can help in the BPM education
  - Can help business analysts to better understand the process models they should design
Challenges related to collections

C20: Discover Model Mapping

C21: Calculate Model Similarity

C22: Search Model

C23: Discover Object Lifecycle

C24: Discover Ontology

C25: Categorize Model

Mendling et al. 2015
Research Plan on NLP for Requirements Engineering

- Extend our approach to support a larger number of BPMN elements
- Filter natural language perspectives such as data and events
- Improve the quality of process descriptions
- Improve research on the extraction of declarative constrains from natural language
References


• Leopold, Pittke, Mendling: Automatic Service Derivation from Business Process Model Repositories via Semantic Technology. Journal of Systems and Software. Accepted for publication.

• Leopold, Mendling, Günther: What we can learn from Quality Issues of BPMN Models from Industry. IEEE Software. Accepted for publication.
