NLP as a Service: An API to Convert between Process Models and Natural Language Text

Thomas Freytag, Benjamin Kanzler, Nils Leger, Daniel Semling[1]

[1]*Cooperative State University (DHBW) Karlsruhe, Germany*

Abstract

An interesting extension of contemporary BPM tools is the ability to export graphical process models into natural language and vice versa, i.e. to mine a graphical process model from a description in natural language. Both transformations are computationally complex Natural Language Processing (NLP) problems and usually require elaborated algorithms, data structures, and IT resources. This contribution describes the architecture of a platform providing NLP-based conversion in both directions *as a service*, i.e. via a public webservice interface, addressing BPM users or modelers as well as software engineers aiming to add NLP features to existing BPM tools.

Keywords

Natural Language Processing, NLP as a Service, Webservice, BPMN, PNML

1. Introduction

Graphical modeling languages allow stakeholders to collectively describe executable business processes on a conceptual level. However, these models frequently co-exist with informal process descriptions specified in natural language. This has brought up the question in BPM research how to transform both representations into each other, i.e. to automatically generate the textual equivalent of a given graphical model (e.g. discussed in [4, 5]) or to automatically generate an equivalent graphical model from a given textual description (e.g. suggested in [3, 7]). Various practical use cases are obvious for this approach:

- *Documentation*: Automated creation of a textual process descriptions from models (e.g. for QM requirements)
- *Visualization:* Automated mining of graphical process models from textual descriptions (e.g. adding models to a process handbook)
- *Comformance*: Comparison between "to-be" textual and "as-is" graphical process descriptions
- *Inclusion:* Enabling visually impaired individuals to work in BPM teams by providing a character-based modelling interface

Proceedings of the Demonstration & Resources Track, Best BPM Dissertation Award, and Doctoral Consortium at BPM 2021 co-located with the 19th International Conference on Business Process Management, BPM 2021, Rome, Italy, September 6-10, 2021

thomas.freytag@dhbw-karlsruhe.de [kanzler.benjamin | leger.nils2 | semling.daniel]@edu.dhbw.karlsruhe.de (T. F. B. K. N. L. D. Semling)

This contribution is about a resource allowing NLP-based transformations between graphical models and natural language via a public webservice. Currently supported modeling languages are PNML (workflow Petri nets) and BPMN. The resource can either be used on a publicly available shared server or tailored, built, and deployed on own premises.

2. Architecture of NLP as a Service

The general architecture of the resource is shown in figure 1. The two webservice applications for verbalizing a model into a text ("Process2Text" or short P2T) and for mining a model from a text ("Text2Process" or short T2P) are deployed in separate docker containers[1], providing all necessary NLP algorithms and data structures (named here "NLP libraries"). The algorithms of these two components are described in more detail in the following sections 3 and 4 of this paper.

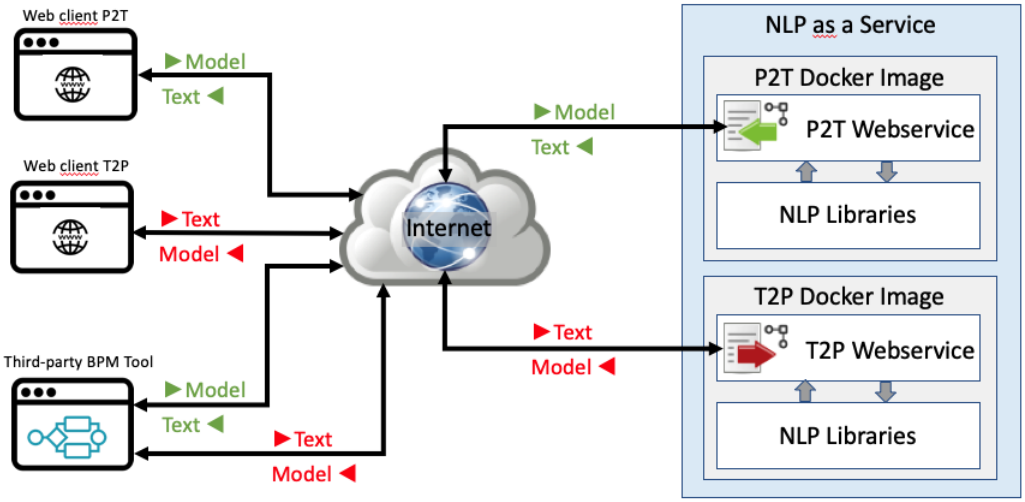

Figure 1: Architecture of NLPaaS

On the frontend side, users can access the provided NLP functions via a simple website or by calling the REST API directly on source code level. Both functions accept a string parameter as input and return a string as output. Examples scenarios for using the service from the client perspective are shown in section 5. More details on the structure of the webservice API and how to call the service from other software products can be taken from the additional document delivered with this publication ([3]).

3. Backend Software of the P2T Webservice

We first have a look at the P2T part of the application, the T2P part will follow in the next section. The algorithm is mainly based on the original open source software developed belonging to the

[1]Docker is an open-source container virtualization software (http://www.docker.com)

research performed by Henrik Leopold et. al. [5] and has been adapted and extended over the years by the authors of this paper. Figure 2 shows the outline of the P2T algorithm.

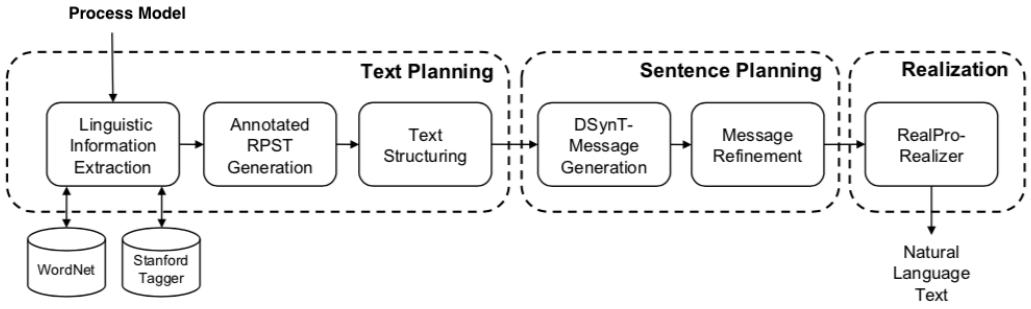

Figure 2: The Process2Text algorithm (adapted from [4])

The conversion takes a process model as input and starts with extracting the relevant linguistic information ("Text Planning"), mainly by analyzing labels of tasks and task-related resource assignments. The tagging and parsing is done with support of the freely available Stanford CoreNLP toolset [2] and the linguistic database WordNet[3]. The result is converted to a Refined Process Structure Tree (RPST), a generic data structure for process graph components as introduced in [6]. Afterwards various heuristics to restructure and fine-tune the RPST ("Text Structuring) are applied, followed by "Sentence Planning" in order to create grammatically correct sentence structures. The RPST leaves are processed and enriched with semantic information derived from the control flow of the graph, leading to the construction of a Deep Syntactic Tree (DSyncT). After the additional refinement of text fragments, a realizer tool is applied in order to generate a naturally-sounding English text which is finally returned to the caller. The source code of the service and additional documention can be found on GitHub under http://github.com/tfreytag/P2T.

4. Backend Software of the T2P Webservice

The opposite direction of the transformation (T2P) is as well based on an existing open source software product implementing the research published in [3] and [7] and has been enhanced and adapted. As in the case of P2T, the algorithms make use of third-party open source software tools and data structures. Figure 3 shows the major steps of the T2P algorithm. As a first step, the input string is analyzed by the Stanford CoreNLP parser and split up into sentences and words according to the underlying grammar rules. The subsequent steps make use of WordNet (see above) and FrameNet[4]. After this, the semantically relevant elements of the textual process model (i.e. actors, actions, business objects, resources) can be determined. The result is written into a data structure "WorldModel" containing all contributing model elements and their interrelations. Finally, the WorldModel is converted into the preselected native process

[2]http://nlp.stanford.edu/software/lex-parser.shtml
[3]http://wordnet.princeton.edu
[4]http://framenet.icsi.berkeley.edu/fndrupal

model format. The last step is to compute a visual layout for the automatically generated process graph. This can be either done on server side or on client side. The source code of the service and additional documention can be found on GitHub under http://github.com/tfreytag/T2P.

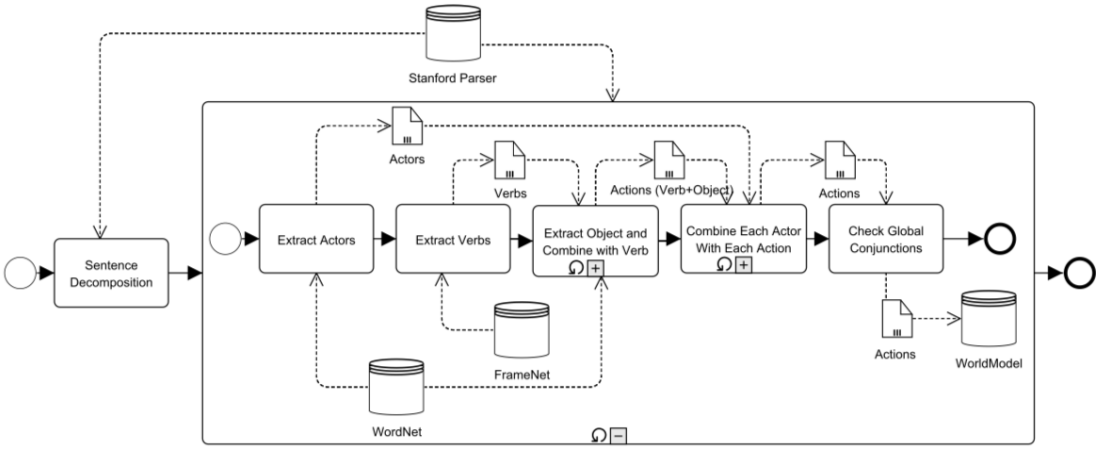

Figure 3: The Text2Process Algorithm (taken from [3])

5. Client Software using NLP as a Service

A reference software for using NLP as a Service is WoPeD[5], a tool which has already been presented at previous BPM conferences [1, 2].

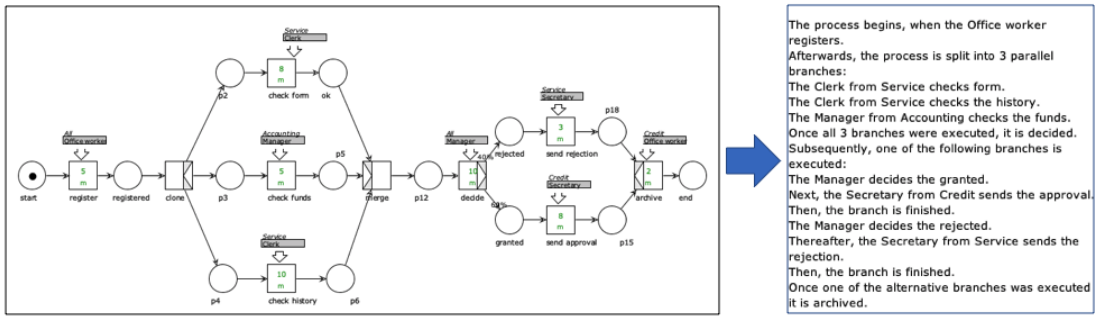

Figure 4: Example transformation from process to text (with Petri net from WoPeD)

Figure 4 shows how a graphical model (a loan application) created in WoPeD has been successfully converted into natural language via the P2T webservice, figure 5 illustrates the opposite direction (T2P). Additionally, both webservices can be also accessed from a simple web interface. More details on how to set up and make use of these interfaces is described in the document provided at [8].

[5]Workflow Petrinet Designer, http://www.woped.org

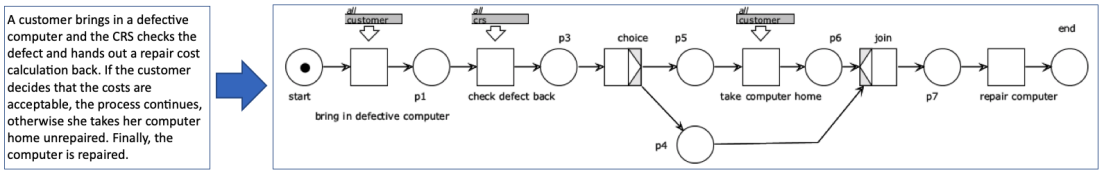

Figure 5: Example transformation from text to process (with Petri net from WoPeD)

6. Conclusion and Outlook

The described resource "NLP as a Service" provides a standardized interface to access P2T and T2P conversion features for BPM tools. Modelers can either export their process model into a readable text, or create a process model from a given textual description. Various practical as well as scientific use cases can be identified. An existing reference application is WoPeD [2]. In the future, further enhancements of both NLP algorithms as well as support for other modelling languages like e.g. EPCs are planned.

References

[1] Freytag, T.; Allgaier, P.; Burattin, A.; Danek-Bulius, A.: WoPeD – A "Proof-of-Concept" Platform for Experimental BPM Research Projects, In: Proceedings of the BPM Demo Sessions, Barcelona, September 2017.

[2] Freytag, T.; Allgaier, P.: WoPeD goes NLP: Conversion between Workflow Nets and Natural Language. BPM Conference Demo Session. http://ceur-ws.org/Vol-2196/BPM_2018_paper_21.pdf, Sydney, Australia, 2018.

[3] Friedrich, F.; Mendling, J.; Puhlmann, F.: Process Model Generation from Natural Language Text. In: Mouratidis H.; Rolland C. (eds): Advanced Information Systems Engineering (CAiSE 2011). LNCS vol. 6741, Springer, Berlin, 2011.

[4] Leopold, H.: Natural Language in Business Process Models: Theoretical Foundations, Techniques, and Applications. Lecture Notes in Business Information Processing, Vol. 168, Springer, Berlin, 2013.

[5] Leopold, H.; Mendling, J.; Polyvyanyy, A.: Generating Natural Language Texts from Business Process Models. In: Ralyté, J., Franch, X., Brinkkemper, S., Wrycza, S. (eds) Advanced Information Systems Engineering. CAiSE 2012. LNCS vol. 7328. Springer, Berlin, 2012.

[6] Polyvyanyy A.; Vanhatalo J.; Völzer H.: Simplified Computation and Generalization of the Refined Process Structure Tree. In: Bravetti M.; Bultan T. (eds) Web Services and Formal Methods. LNCS vol. 6551, Springer, Berlin, 2010.

[7] Riefer, M.; Ternis, S.; Thaler, T: Mining Process Models from Natural Language Text: A State-of-the-Art Analysis. Processings of Multikonferenz Wirtschaftsinformatik (MKWI 16), March 9-11, Illmenau, Germany, 2016.

[8] Freytag, T.; Kanzler, B.; Leger, N.; Semling, D.: NLP as a Service - Technical Documentation, https://woped.dhbw-karlsruhe.de/wp-content/uploads/TechnicalDocumentation.pdf. Accessed on July 30th, 2021.