

Natural Language Processing for Business Process Analysis

Han van der Aa¹, Henrik Leopold^{2,3}, Kiran Busch², Adrian Rebmann¹

¹ University of Mannheim, Mannheim, Germany,
{han.van.der.aa|rebmann}@uni-mannheim.de

² Kühne Logistics University, Hamburg, Germany,
{henrik.leopold|kiran.busch}@the-klu.org

³ Hasso Plattner Institute, Potsdam, Germany

Abstract. Natural Language Processing (NLP) has become an essential tool for many organizations aiming to analyze and understand the vast amounts of text data they generate. The latest developments related to language models have significantly boosted the analytical capabilities of NLP tools and have created completely new use cases. In this tutorial, we will focus on the intersection of NLP and Business Process Management (BPM) and explore how NLP can support various BPM analysis tasks. We first introduce fundamentals of NLP and explore how language models work. Then, we focus on the automated analysis of textual descriptions, after which we turn to the analysis of process-oriented artifacts, where we show how NLP can be used to obtain novel insights from process models and event logs. These parts are followed by a hands-on exercise session, in which participants will learn how to use general and process-specific NLP libraries and techniques. Finally, we conclude the tutorial with a discussion of future directions. After the tutorial, participants will have learned about the fundamentals of NLP, the potential of using NLP in the context of BPM, and how to apply NLP to their own BPM research and analyses.

1 Content outline

The content of this tutorial will consist of the following parts:

Introduction to NLP and Language Models. (15 minutes)

In this part of the tutorial, we will provide the audience with the prerequisite knowledge necessary to understand state-of-the-art NLP tools. Besides introducing basic concepts, such as word embeddings, we will provide a gentle introduction to language models. Afterwards, we explain the core idea behind the transformer architecture and show they can be easily integrated using Python.

Extracting Process Information from Texts. (15 minutes)

This part of the tutorial will focus on the extraction of process information from textual process descriptions. We will show the challenges involved in this task, such as the large degree of variety, describe how

this task has so far been tackled, and briefly demonstrate how the extracted information can be used for downstream tasks such as process model extraction [3] and conformance checking [1].

Using NLP on Process-oriented Data. (15 minutes)

Next, we turn to the use of NLP for the analysis of process-oriented artifacts, focusing on process models and event logs. We will show how NLP can be used to extract semantic information from these artifacts, such as the actions and business objects contained in them. Subsequently, we will show how this extracted information can be used for purposes such as anomaly detection [2] and event abstraction [4], bringing a new layer of insight to process mining and analysis.

Hands-on session. (30 minutes)

In this part, we will provide the attendees with practical demonstrations and exercises to extract and analyze process information from both textual descriptions and real-world event logs. To support this part, we will establish dedicated Jupyter notebooks that guide the attendees through the different steps, so that they can quickly gain insights. The notebooks will be designed to encourage audience members to try out different prompts (for ChatGPT) and also apply them on their own examples. During the exercises, we will provide guidance and answer any questions that may arise. To participate, audience members just need a laptop with Python, for which we will provide installation instructions.

Future of NLP Applications in BPM. (15 minutes)

In the final part of the tutorial, we will discuss how we expect that the role of NLP will further develop in the context of BPM in the future and engage the audience in a discussion on this matter.

2 Intended audience

Our tutorial targets academics, practitioners, and students who are interested in learning about the application and potential of NLP for BPM and process mining. Audience members are not expected to have any prior experience with NLP. However we do expect them to be familiar with the fundamentals of BPM and process mining.

References

1. Van der Aa, H., Leopold, H., Reijers, H.A.: Checking process compliance against natural language specifications using behavioral spaces. *Information Systems* **78**, 83–95 (2018)
2. Van der Aa, H., Rebmann, A., Leopold, H.: Natural language-based detection of semantic execution anomalies in event logs. *Information Systems* **102**, 101824 (2021)
3. Bellan, P., Dragoni, M., Ghidini, C.: Process extraction from text: state of the art and challenges for the future. *arXiv preprint arXiv:2110.03754* (2021)
4. Rebmann, A., van der Aa, H.: Enabling semantics-aware process mining through the automatic annotation of event logs. *Information Systems* **110**, 102111 (2022)