Chair of Information Systems and Business Process Management (i17)
Department of Computer Science
TUM School of Computation, Information and Technology
Technical University of Munich

TUM

Proposal to Bachelor's thesis

Auto-generation of business process models using natural language processing

Shuaiwei Yu

Thesis for the Attainment of the Degree
Bachelor of Science

at the TUM School of Computation, Information and Technology,
Department of Computer Science,
Chair of Information Systems and Business Process Management (i17)

Examiner
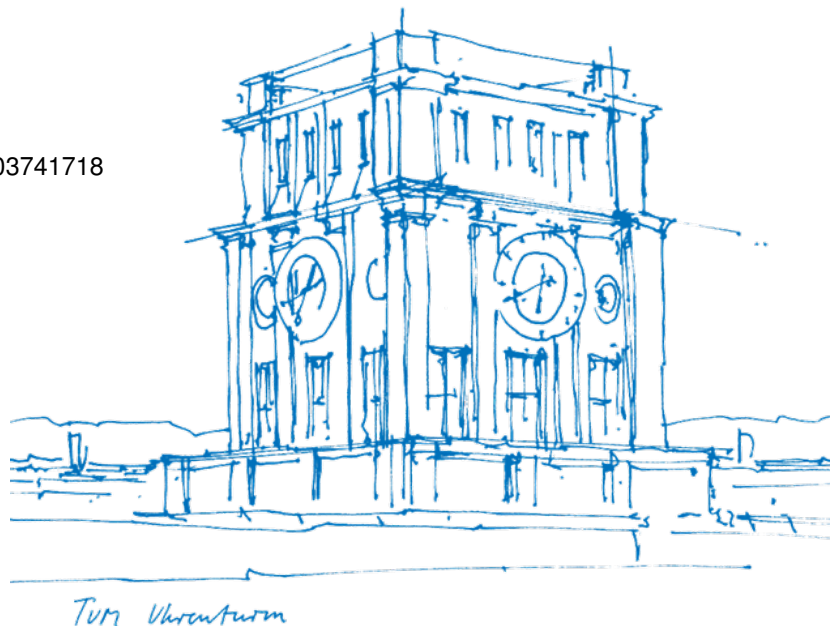Prof. Dr. Stefanie Rinderle-Ma

Supervised by
Catherine Sai, M. Sc.

Submitted by
Shuaiwei Yu
Matriculation Number: 03741718

Submitted on
15.04.2023

TUM Uhrenturm

Abstract

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Keywords: *Nature language processing, business process models, Artificial intelligence.*

Contents

List of Tables

List of Figures

Motivation

Business processes are fundamental elements for companies and organizations. They aggregate all the tasks, activities, and timelines involved in companies' workflow whose aim is to provide business or to create value [1]. Business Process Modeling Notation, also known as BPMN is a modeling language describing such workflows by using graphical notations and thus provides an easily understandable overview of the operations performed in the organization for all business users [3].

Due to the importance of Business processes, leveraging the BPMN techniques can positively affect an organization's performance and thus increase its productivity. However, not everyone is familiar with the BPMN designing techniques. Consequently, managers, along with other process participants prefer using natural language to define business processes. As a result, organizations usually have a large amount of information stored as text documents [1]. There is a need to translate text documents into the process model regarding such a situation. However, process modeling is not a simple task, but is time-consuming and experts with professional knowledge are required.

Over the past years, the development of AI techniques brought solutions to many technical difficulties. Natural Language Processing (NLP), as one of the AI's branches, could possibly address the problem of the difficulties in process modeling. Natural Language Processing is an interdisciplinary discipline focusing on the study of algorithms that enable the computer to understand and process the human language[4]. During the understanding and processing of the natural language text, NLP performs three types of analysis: Firstly, morphological analysis is performed, which analyze the structure of words. The syntactic analysis then explores the grammar relationship between words in sentences, deciding which grammar category the word belongs to. Finally, semantic analysis is executed, which leverages the afore analyses to define the meaning of the text based on the knowledge of sentence structure and the relationship between words [1].

The unique features of the NLP technique make it very suitable for exploiting information from the text documents that record the firm's business process and then analyzing the data to generate the process models automatically. This paper serves as a proposal to suggest using NLP

to extract the information from text written in nature language and automatically generate the corresponding business model.

Research Questions

The main research question (**RQ**) is formulated as: "*How can business process models be generated from regulatory documents automatically using the Nature language processing technique?*". To better answer the main research question, three embedded aspects can be revealed: **RQ1**: "*which NLP methods can be used to extract information?*"; **RQ2**: "*How can the extracted information be analyzed and composed to generate business process models?*" and finally **RQ3**: "*How does the proposed approach perform with different kinds of input documents?*"

Currently, there exist various tools, libraries, and dependencies for NLP. Therefore, the first research question **RQ1** tries to figure out which methods are the most suitable ones to use to extract information from regulatory documents. The methods should be able to separate sentences, label each word in a sentence with corresponding syntactical tags, and analyze the grammatical relationships between words. By doing so, we are able to explore the information hidden behind the natural language and thus use them for further operation. In the next step, **RQ2** explores how to use syntactical and grammatical information to determine events of business processes, identify the conditional restraints ("and" or "or"), and the sequential orders of business processes. Once such information is acquired, an algorithm should be developed to combine all the business processes in a logical order. In the end, the composed process model should be well visualized. The last research question **RQ3** tries to discover the adaptability of the proposed model: How well does the method perform with the document other than the regulatory document? Does the accuracy of the outcome decrease with other kinds of documents? A sets of different input documents will be prepared and a corresponding benchmark will be performed.

Research Methodology

Related Work

In order to learn the current state-of-the-art methods of auto-generating business process models and thus answer the research question comprehensively, a systematic literature review must

Table 1

Overview of Systematic Literature Review Protocol

Database	hits	selected	
IEEE	56	0	
Springer	275	0	
ACM	201	0	
Google scholar	0	0	
Result horizontal search	0	0	
Vertical search	0	0	add papers
Overall	0	0	

be performed so that we can learn what kind of efforts are made as well as what are the most preferred techniques and what open challenges exist. The literature review is conducted under the guidance of Kitchenham et al. given in their paper [2]. The work consists of several stages: Firstly, the electronic database used to run the search is chosen. Then the selection criteria are defined, and articles are filtered accordingly. After that, a horizontal search will be run to cover as many papers as possible. Finally, a list of the final literature is studied carefully, and helpful information is extracted.

In order to perform a comprehensive literature review, we chose three most famous electronic databases, i.e., IEEE, Springer, and ACM. Nevertheless, only using these three databases, There is still a minor chance that some important articles will be missed. Therefore, we also used Google scholar as a complement because it covers a wide range of literature, from conference papers to degree theses. The search string used for the literature review is developed using two keywords, which are the most important ones for our research: *business process model* and *natural language processing*.

In the next step, inclusion and exclusion criteria should be defined so that the articles

Method/Approach (theoretical)

Application (practical)

Evaluation

Time Plan

Bibliography

[1]   Ana Cláudia de Almeida Bordignon et al. "Natural language processing in business process identification and modeling: a systematic literature review". In: *Proceedings of the XIV Brazilian Symposium on Information Systems*. 2018, pp. 1–8.

[2]   Barbara Kitchenham. "Procedures for performing systematic reviews". In: *Keele, UK, Keele University* 33.2004 (2004), pp. 1–26.

[3]   Bilal Maqbool et al. "A comprehensive investigation of BPMN models generation from textual requirements—techniques, tools and trends". In: *Information Science and Applications 2018: ICISA 2018*. Springer. 2019, pp. 543–557.

[4]   Konstantinos Sintoris and Kostas Vergidis. "Extracting business process models using natural language processing (NLP) techniques". In: *2017 IEEE 19th conference on business informatics (CBI)*. Vol. 1. IEEE. 2017, pp. 135–139.

Appendix

Table 2
Your first table

Value 1	Value 2	Value 3
α	β	γ
1	1110.1	a
2	10.1	b
3	23.113231	c

A note describing the table.

Figure 1

My Figure Caption

A note describing the figure

Declaration of Academic Integrity

I hereby declare that the thesis submitted is my own unaided work. All direct or indirect sources used are acknowledged as references.

I am aware that the thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the thesis as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of my work with existing sources I agree that it shall be entered in a database where it shall also remain after examination, to enable comparison with future theses submitted. Further rights of reproduction and usage, however, are not granted here.

This thesis was not previously presented to another examination board and has not been published.

Garching, 15.04.2023   Shuaiwei Yu