NLP4BPM - Natural Language Processing Tools for Business Process Management

Luis Delicado, Josep Sànchez-Ferreres, Josep Carmona, and Lluís Padró

Computer Science Department
Universitat Politècnica de Catalunya
Barcelona, Spain
{ldelicado,jsanchezf,jcarmona,padro}@cs.upc.edu

Abstract. Business Process Management is facing a drift in the way process information is used within an organization. To reach a wide audience, organizations keep parallel representations of process information, thus making the processes to be understandable by everyone. However, this poses a challenge on the synchronization and transformation between different process representations. NLP4BPM is an environment to support such crucial tasks. It combines Natural Language Processing tools with process-oriented techniques, and through its web-based interface can be easily accessed from any device.

1 Introduction

Process documentation is a key aspect of Business Process Management. Nowadays organizations store processes descriptions in various representations. The reason for this is the different nature stakeholders have: while textual descriptions of processes are well-suited for non-technical members, they are less appropriate for describing precise aspects of the underlying process [1]. In contrast, formal and graphical process notations (e.g., BPMN) are unambiguous representations which can be the basis for automating the corresponding processes within the organization [2], but they are oriented to specialized members. In this context, due to the evolving nature of processes, there is a high risk of having deviations between the different representations, a problem that may have serious consequences for any organization [3].

In the last decade, the field of *Natural Language Processing* (NLP) has become mature enough to provide text analysis accuracies accurate enough for a variety of applications. Currently, there are several powerful open-source libraries [4,5,6,7] that can be easily integrated into any software project, turning linguistic analysis into a commodity in application contexts such as BPM.

There are several approaches one might consider in order to automate dealing with consistency issues in multiple representations. One approach is to generate a textual description in natural language using the most up-to-date process model. There is also the complementary technique, converting the textual description into a formal process model. The third approach consists on performing a comparison of two existing process representations.

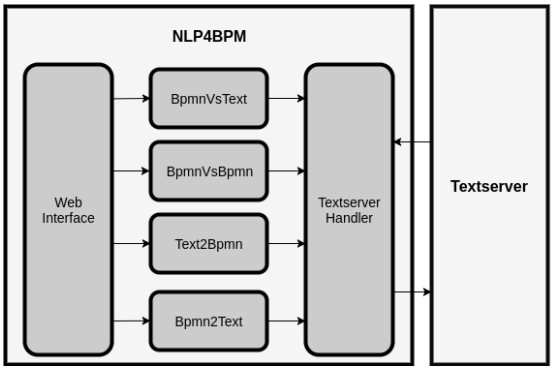

Fig. 1. Overview of the project architecture

In this paper, we introduce NLP4BPM, an online platform offering NLP based services for the aforementioned BPM use-cases: Converting between several process model representations as well as comparing process models between themselves and to textual descriptions. The interface presented covers adaptations or modifications of techniques that have appeared in the last years [1,8,9].

2 Tool Description

The tool we present offers several functionalities in the context of semantic understanding of business processes, all of which use natural language processing techniques:

Text2BPMN: Converts a textual description into a BPMN process model.
BPMN2Text: Converts a BPMN process model into a textual description.
BPMNvsText: Computes an *alignment* between a model and a textual description in order to compare them. An alingment is a mapping between tasks in the model and sentences in the text. The implementation follows the technique presented in [9]
BPMNvsBPMN: Computes an alignment between two different BPMN models. Here the alignment is between the set of tasks of the two process models. The technique implemented in this module are based on Relaxation Labeling algorithms.

The different techniques are developed as independent modules and unified under the same graphical interface, which is presented in a user-friendly manner. A screencast of the tool can be found at `https://youtu.be/mAsXnEBQMx0`

2.1 Architecture and Maturity

An overview of the architecture of the project can be seen in Figure 1. The core of the application are the four main modules, implemented as separate

Java projects. The graphical interface is build as a JavaServer Faces (JSF) web application which integrates the four main modules, and provides a global configuration interface to all of them. By using the JSF framework, we obtain a robust way to integrate Java code into a responsive web application that can be accessed from all kinds of devices: From personal computers to smartphones.

At the very core of all four techniques, there is natural language processing. In order to offer a scalable, robust system, we forward the NLP tasks to *TextServer*[1], a distributed REST-like web service running FreeLing [5] underneath[2]. The communication between any of the modules and TextServer is through a JSON API.

The tool is currently under heavy development with the four main modules being improved. The comparison modules: BPMNvsBPMN and BPMNvsText are currently at a stable state, while the translation ones: BPMN2Text and Text2BPMN are early prototypes still being improved.

2.2 Usage and Example

The NLP4BPM web application is available from `nlp4bpm.cs.upc.edu`. After login, a main menu is shown to the user with four tabs corresponding to each of the described modules, as seen in Figure 2.

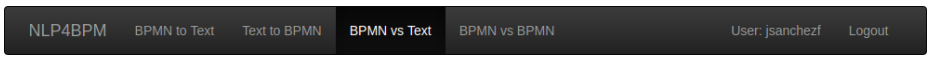

Fig. 2. The four main modules, shown on the application's navigation bar

The left menu contains the configuration parameters of each algorithm. The configurations vary for each module, but all share a set of common parameters:

Language Due to FreeLing's multi-language support, our tool also provides support for performing tasks on many languages besides English, among which we can find Spanish, French or German. Note that this feature is under construction for some of the modules.

Input Files The supported input files are BPMN XML files for the process models and raw text files for the textual descriptions. Additional format support is planned as a future extension for several rich document formats as well as different process model notations. Also, the support for the (recently IEEE standard) XES event logs format will be added soon. The users can upload their own input files or use one of the provided examples.

The four modules offer a very similar interface. Once the user has set the module-specific parameters and specified the textual description and/or BPMN model, the *Execute* button runs the corresponding algorithm.

[1] `textserver.cs.upc.edu/textserver`
[2] `nlp.cs.upc.edu/freeling`

In the translation modules, the user is shown the resulting BPMN model or textual description, which is saved in the user's online directory and can be downloaded. In the comparison models, a graphical representation of the comparison is shown to the user.

Figure 3 shows an example of the resulting comparison from the BPMNvsText module. A BPMN process model is shown with its tasks and gateways highlighted in colors. The colors code is used on the sentences below, in order to show the alignment. For additional information, the user can click one of the matchings to get information about which linguistic features were found in common for the sentence and the task.

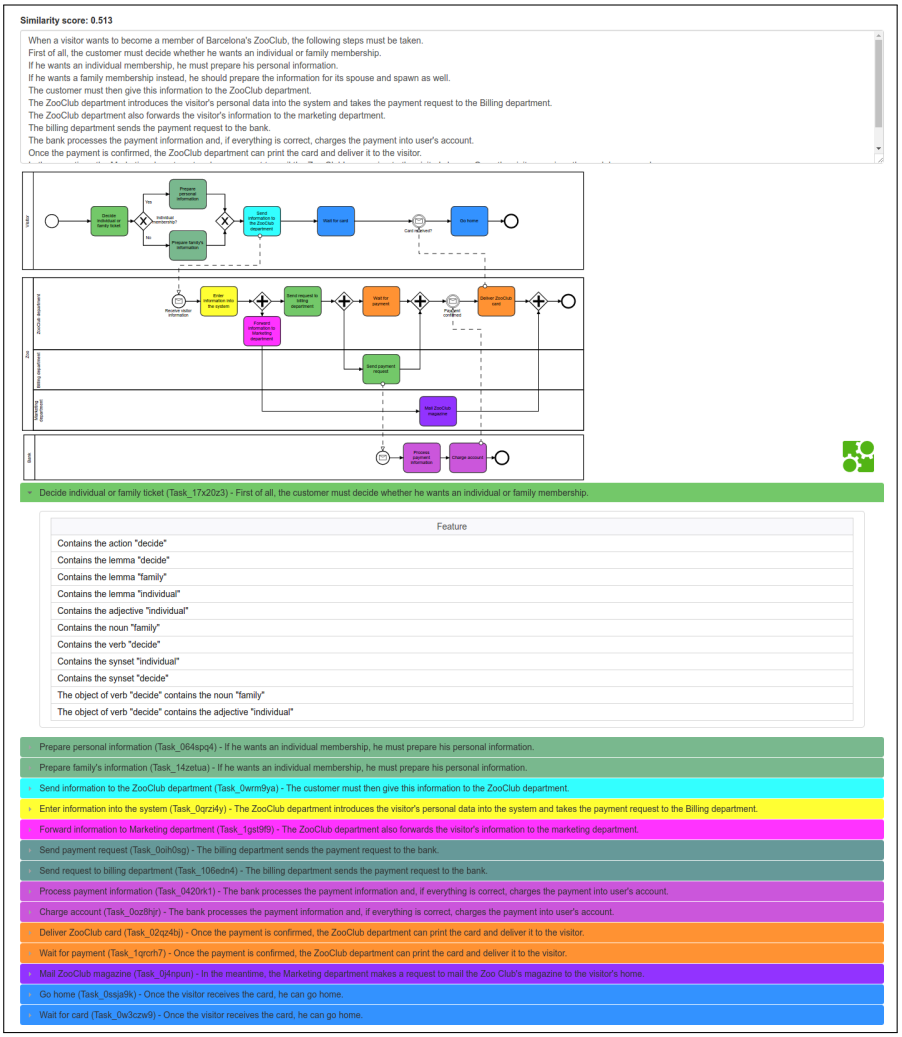

Fig. 3. Execution example for the BPMNvsText module

3 Significance for the BPM Field

By providing an easy-to-use environment for the transformation and comparison of process descriptions, organizations can reach an unprecedented process awareness. This may be crucial to avoid wrong expectations on the processes within a company, but also to ensure all the important agents running these processes are aligned.

We also foresee interesting applications of the environment proposed. For instance, the environment can be used to assist in the process of process modeling [10].

References

1. Leopold, H., Mendling, J., Polyvyanyy, A.: Supporting process model validation through natural language generation. IEEE Trans. Software Eng. **40**(8) (2014) 818–840
2. Dumas, M., Rosa, M.L., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management. Springer (2013)
3. van der Aa, H., Leopold, H., Mannhardt, F., Reijers, H.A.: On the fragmentation of process information: Challenges, solutions, and outlook. In: Enterprise, Business-Process and Information Systems Modeling - 16th International Conference, BPMDS 2015, 20th International Conference, EMMSAD 2015, Held at CAiSE 2015, Stockholm, Sweden (June 2015) 3–18
4. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D.: The Stanford CoreNLP natural language processing toolkit. In: Association for Computational Linguistics (ACL) System Demonstrations. (2014) 55–60
5. Padró, L., Stanilovsky, E.: Freeling 3.0: Towards wider multilinguality. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC, Istanbul, Turkey (May 2012) 2473–2479
6. Bird, S., Loper, E., Ewan, K.: Natural Language Processing with Python. O'Reilly Media Inc (2009)
7. Apache Software Foundation: Apache OpenNLP. `opennlp.apache.org` (2010)
8. Friedrich, F., Mendling, J., Puhlmann, F.: Process model generation from natural language text. In: Advanced Information Systems Engineering - 23rd International Conference, CAiSE 2011, London, UK, June 20-24, 2011. Proceedings. (2011) 482–496
9. Sànchez-Ferreres, J., Carmona, J., Padró, L.: Aligning textual and graphical descriptions of processes through ILP techniques. In: Advanced Information Systems Engineering - 29th International Conference, CAiSE 2017, Essen, Germany, June 12-16, 2017, Proceedings. (2017) 413–427
10. Weber, B.: Investigating the process of process modeling: Towards an in-depth understanding of how process models are created. In: 18th IEEE International Enterprise Distributed Object Computing Conference, EDOC 2014, Ulm, Germany, September 1-5, 2014. (2014) 180