Management School – University of Liège

Irrelevant Sentences Detection for Automated Business Process Modeling

Name: Julie Jamar

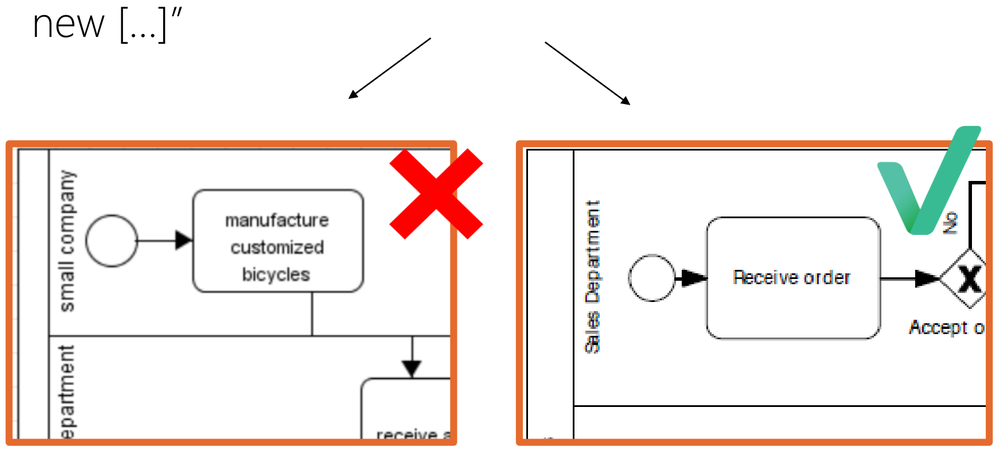
Introduction

In Business Process Management, a full automation of the modeling process can reduce the resource requirements by up to 60%. Current work propose solutions only working when process descriptions are sequential and do not contain noise like irrelevant information. Because process descriptions are unstructured, described on a meta level, it may contain both relevant and irrelevant information; parts of the textual descriptions might be irrelevant for the generated process model, and it is important to identify and ignore them. Interestingly, no study addressing this problem was found, although this provide the starting point for the construction of a modeling engine.

Problem illustration

The following example illustrates the issue of relevance in process modeling:

"*A small company manufactures customized bicycles*. Whenever the sales department receives an order, a new [...]"



Effective filtering technique is needed to reliably identify and ignore certain parts of the textual description. However, some nuances have to be considered before cancelling parts of the text:

	Must be modeled	Must not be modeled
Bring information for modeling	Relevant sentence	Informative sentence
Do not bring information for modeling	N/A	Irrelevant sentence

Table 1: Existing nuances in sentence relevance for Business Process Modeling.

Generalization

Irrelevant Sentences generally comes with:

1 Introductory Sentences

Descriptions sometimes begin with a sentence providing very general information on the context.

→ "*The examination process can be summarized as follows*. The process starts when [...]"

2 Example Sentences

Examples are used to create a more vivid text, but are unwanted in a generalized representation.

→ "[...] a chance to edit it, *for example to correct errors or better describe an expense*."

3 Detail Information

Textual descriptions sometimes contain additional details, enabling a better understanding.

→ "[...] select the underwriters. *Underwriters act as financial midwives to a new issue*. Usually they [...]"

TF-iDF for Sentence Weighting

Text summarization algorithms represent a promising basis for identifying Irrelevant Sentences in cases **1** and **3**. We based our first experiments on a TF-IDF sentence weighting algorithm initially created for text summarization.

First results are really satisfying; Irrelevant Sentences frequently get remarkable scores, leading to the conclusion that this is an promising path to explore.

Keywords Search

Case **2** could be easily solved using keywords search, with "*for example*", "*for instance*", "*i.e.*" and so on.

Conclusion & Discussion

Text summarization algorithms represent a promising basis for identifying Irrelevant Sentences, but they still need adaptations to be more effective.

Future work:

- Insertion of coefficients related to Irrelevant Sentences features in the sentence weighting algorithm in order to increase its performance.
- Approach from another angle: identifying Relevant Sentences and then deduct the Irrelevant ones by elimination.

Julie Jamar
julie.jamar@uliege.be
PhD Candidate | QuantOM - Centre for Quantitative Methods and Operations Management
HEC - Management School
Rue Louvrex 14, 4000 Liège
BELGIUM