

## Doctoral meeting: 'Process-to-Text: A Framework for the Description of Processes in Natural Language'

**Data:** venres, 30 abril, 2021 -  
10:00

**Lugar:** Virtual event

**Poñente(s):** Yago Fontenla Seco (CiTIUS PhD  
Candidate)

**Idioma:** Inglés

**Streaming:** Este evento emitirase en directo. A ligazón á emisión poñerase aquí cando esté dispoñible.



Processes constitute a useful way of representing and structuring the activities and resources involved in organization Information Systems. Daily, more event data is being produced and recorded, and process mining goal is to exploit these data, by automatically discovering the underlying process model, to extract valuable, process related information in a meaningful way that can be used to understand what is really happening in a process.

Process models represent in a graphical manner the activities that take place in a process as well as the dependencies among them. They tend to be enhanced with properties such as temporal information, process execution-related statistics, etc. This information is shown to users through visual analytic techniques, however, in real scenarios process models are highly complex and the number of properties that can be added to the process model is very high. This makes process models and visual analytics nearly impossible to be interpreted and understood by users, as deep knowledge of process modeling and analytics is required.

Natural Language Generation (NLG) and Linguistic Descriptions of Data (LDD) fields aim to provide users with textual descriptions that summarize the most relevant information of some data that is being described. Research suggests that in some domains knowledge and expertise are required to understand graphical information and proves that specialists can take better decisions based on textual summaries than on graphical displays.

Therefore, natural language descriptions seem a good approach to enable or enhance the understanding of processes and its analytics as they can summarize, combine, and communicate information in ways it would not be possible with visual representations.

The objectives of this thesis are then the following:

- To define a formal model able to generate natural language descriptions of both qualitative and quantitative information about processes; including fuzzy temporal and causal information from the process and its attributes, quantifying attributes in time during process life-span and able to recall causal relations and temporal distances between events, among other features.
- Once this model is defined, a complete NLG pipeline will be implemented using this model as its content determination element for choosing which information to communicate and providing an intermediate representation of the information between the original process and the final natural language description.
- Finally, this NLG system will be incorporated over a service-oriented-architecture tool for the interactive generation of process reports using natural language descriptions of processes as well as classical data visualization techniques.

