



POLITECNICO
DI TORINO

SmartData@PoliTO



November 2nd, 2020, 5:30 PM CEST

SmartTalk: on Microsoft Teams



Moreno La Quatra

Politecnico di Torino

Extractive Timeline Summarization by means of Deep Natural Language Processing

ABSTRACT

The large-scale diffusion of the World Wide Web has led to a sudden increase in the amount of published news content. While users benefit from the simple access to relevant content, the abundance of information related to a specific topic could lead to information overload. Timeline summarization aims at processing long streams of news to detect relevant dates and key-insights that better describe the progress of the main event. To address this task, we propose a comprehensive framework that leverages on two main machine learning steps: date selection and text summarization. The former step combines graph modeling and natural language understanding approaches to detect the key dates of the global event. The latter phase leverages on the semantic representation of text to extract concise yet informative summaries of the news published on each selected date. The summarization architecture relies on the fine-tuning of pre-trained deep learning models to estimate single sentence relevance.

BIOGRAPHY

Moreno La Quatra is a PhD student (since 2018) at Politecnico di Torino working in the domain of Multimedia and Text Analysis. His research interests are related (but not restricted to): Natural Language Understanding, Embedding techniques and latent spaces, Data driven content personalization, Multi-modal data integration.