



Contribution ID: 96

Type: **not specified**

From Unstructured Data to Knowledge Discovery: A Novel Deep Learning Framework for Text Mining

Extracting meaningful insights from vast amounts of unstructured textual data presents significant challenges in text mining, particularly when attempting to separate valuable information from noise. This research introduces a novel deep learning framework for text mining that identifies latent structures within comprehensive text corpora. The proposed methodology incorporates an initial sentence classification phase to filter out irrelevant content while preserving essential information. Following this preprocessing step, we implement a deep learning-powered Named Entity Recognition (NER) system that uses predefined feature extraction to identify and extract critical entities, transforming them into structured data formats. We validate our approach using two datasets: BioCreative II Gene Mention (BC2GM) to compare it with other established approaches, and shipping industry datasets—a real-world dataset that contains emails for orders that have been executed. The findings demonstrate that deep learning significantly enhances text mining capabilities, proving its value for extracting essential information from large-scale textual repositories.

Special/ Invited session

Classification

Both methodology and application

Keywords

Text Mining, Deep Learning, Transformers, Feature Extraction, Artificial Intelligence

Primary author: PAPAGEORGIOU, Grigorios (University of Patras)

Co-authors: Mr SKALRATOS, Kyriakos (University of Piraeus); ECONOMOU, Polychronis (University of Patras)

Presenter: PAPAGEORGIOU, Grigorios (University of Patras)

Track Classification: Machine Learning