DOI: https://doi.org/10.36094/sc.v92.2026.An_NLP_Framework_for_Detecting_Abstract.Das.667

ARTICLE

Sci. and Cult. 92 (9–10) : 667-671 (2026)

AN NLP FRAMEWORK FOR DETECTING ABSTRACT-FULL TEXT TOPIC CONSISTENCY IN ACADEMIC PAPERS

AYANTIKA DAS[1], VANDANA YADAV[1] AND SUSHMITA CHAUDHARI[1]*

The abstract of the academic paper is an important component of the academic paper that outlines the contribution and boundaries of the research and findings. However, abstracts are not always a true reflection of the content of the full text and can cause readers, reviewers, editors and indexing systems to assume that the abstract is accurate when it is not. In this paper, a new framework of Natural Language Processing (NLP) has been proposed to automatically identify whether the topic is consistent in the abstract and full text of a research paper or not. It employs the following operations to the framework: the preprocessing, topic extraction, semantic representation, and similarity-based alignment scoring. Latent Dirichlet Allocation (LDA) is a baseline model for probabilistic topic modeling and BERT-based transformer embeddings are a model to capture the contextual semantic similarity. The experiments were run on 1,200 academic papers, with an 80/20 split between training and testing sets, and 5-fold cross-validation, from four benchmark sets: arXiv (300), PubMed (300), ACL Anthology (300), and Semantic Scholar (300). Consistency classification threshold (τ) was chosen through grid search on the validation set, which was 0.75. The proposed guidelines could improve the transparency of research and quality of journal articles for the editors, reviewers, and databases of journals. The field of study known as natural language processing (NLP) is a growing and fast field of research these days, as artificial intelligence (AI) is developing rapidly.