

# Extracting Knowledge from Scientific Literature with an Integrated Text Summarization Approach

Luca D’Aniello<sup>1,4</sup>, Massimo Aria<sup>1,4</sup>, Corrado Cuccurullo<sup>2,4</sup>, Michelangelo Misuraca<sup>3,4</sup>, Maria Spano<sup>1,4</sup>

<sup>1</sup>University of Naples Federico II – luca.daniello@unina.it; massimo.aria@unina.it; maria.spano@unina.it

<sup>2</sup>University of Campania Luigi Vanvitelli – corrado.cuccurullo@unicampania.it

<sup>3</sup>University of Calabria – michelangelo.misuraca@unical.it

<sup>4</sup>K-Synth Academic spin-off – University of Naples Federico II

## Abstract

In recent years, the surge in textual data has led to an information overload, particularly evident in the scientific literature domain due to the rapid pace of document publication. Researchers require tools to manage the sheer volume of publications, enabling the rapid extraction of knowledge. Automatic Text Summarization emerges as a key solution, reducing reading time and streamlining efforts by accentuating key points and generating concise summaries with the most pertinent information. Our innovative approach, named Integrated Text Summarization, aims to enhance extractive algorithms by incorporating specific features of scientific documents—such as keywords and title terms—into the sentence selection process. This method focuses on extracting significant sentences from each section of scientific papers, thereby identifying content that encapsulates the core knowledge of a publication. We will assess the effectiveness of our proposal across various research domains using a selection of scientific papers.

**Keywords:** Automatic Text Summarization; Information extraction; Scientific article features; IMRaD format.