P7. Analyzing Thematic Alignment in Scientific Journals

The core objective is to quantitatively assess whether articles published in a specific journal align with its stated **"Aims & Scope"**. Students develop methodology to model the journal's intended focus and compare it against publication content, potentially identifying thematic drift over time or outlier papers.

1

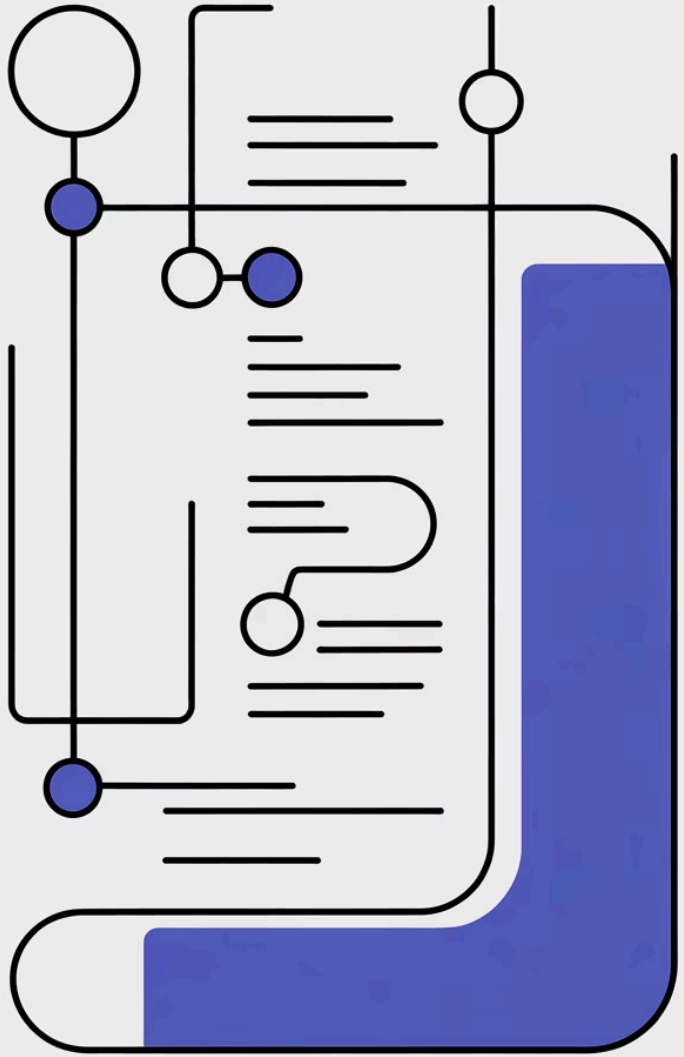
Core Pipeline

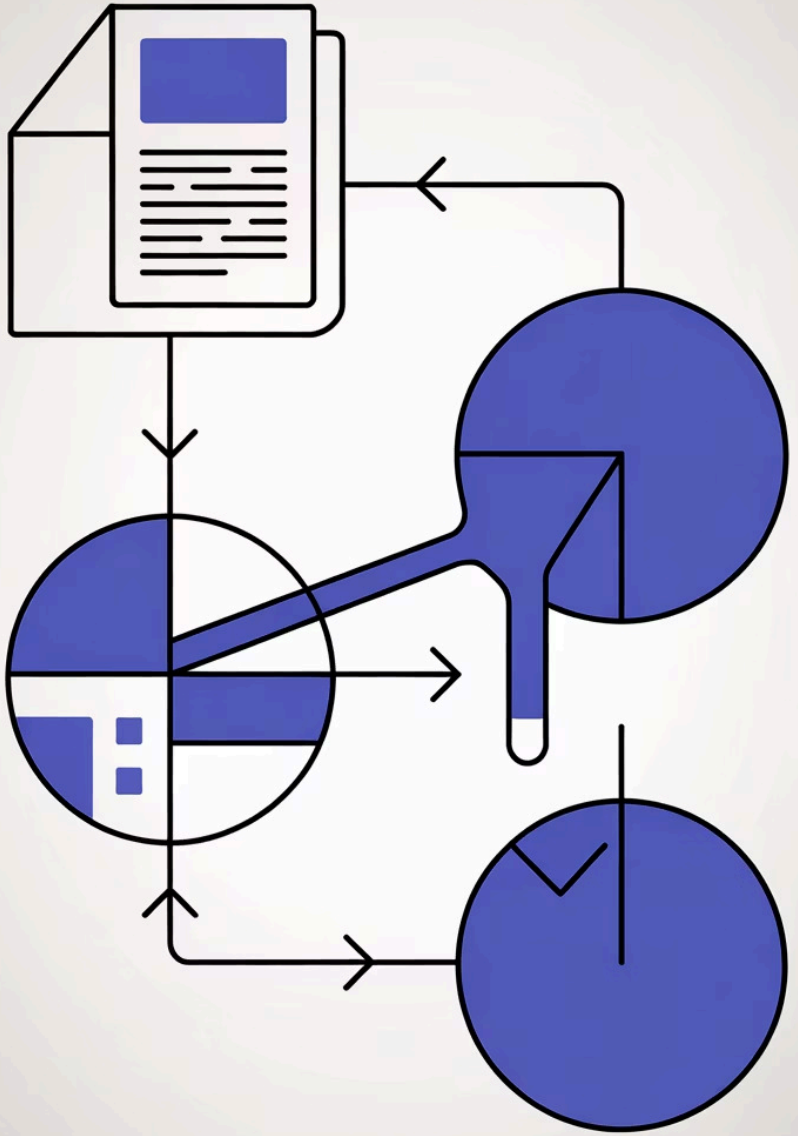
Articles represented as embeddings or topic distributions. "Aims & Scope" serves as thematic reference for alignment scoring.

2

Expected Outcomes

Quantify thematic coherence, identify outlier papers and trends revealing long-term conceptual evolution.





Methodology

Data Curation

Select scientific journal with clearly defined "Aims & Scope" statement as ground truth for intended focus

Content Modeling

Create structured, machine-readable representation capturing core subjects and meaning for both scope and articles

Measure Alignment

Design computational method to measure thematic overlap, generating quantitative alignment scores

Report Findings

Analyze score distribution, visualize results, detect thematic drift, identify outliers with qualitative validation

Dataset & References

Dataset: Use relevant API (arXiv, Semantic Scholar, PubMed) to collect article abstracts from target journal over 5-10 years.

References

- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *EMNLP-IJCNLP*, 3982-3992.
- Picascia, S., et al. (2025). The Atlas of Data Science Research. *IEEE Access*.
- Hassan-Montero, Y., et al. (2014). Graphical interface of the Scimago Journal and Country Rank. *El profesional de la información*, 23(3).