

# SOVEREIGN: To what extent does fine-tuning on BEIR-NL improve R@100 and MRR scores compared to zero-shot baselines for Du

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 29, 2026

## Abstract

Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters). The clustering problem has been addressed in many contexts and by researchers in many disciplines; this reflects its broad appeal and usefulness as one of the steps in exploratory data analysis. However, clustering is a difficult problem combinatorially, and differences in assumptions and contexts in different communities has made the transfer of useful generic concepts and methodologies slow to occur. This paper presents an overview of pattern clustering methods fro

## 1 Introduction

Analysis of: Data clustering. Research goal: To what extent does fine-tuning on BEIR-NL improve R@100 and MRR scores compared to zero-shot baselines for Dutch legal and news domain retrieval?.

## 2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score < 6.5.

## 3 Results

8 papers retrieved. 7 claims extracted, 7 verified. Tribunal: 7.7/10 \$\rightarrow\$ APPROVE (revision\_round=0). Policy: AUTO\_APPROVE.

## 4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

## 5 Extracted Claims

Claim	Verified	Confidence
Clustering is the unsupervised classification of patterns into groups	✓	0.20
The clustering problem has been addressed in many contexts and by researchers in many disciplines	✓	0.27
clustering is a difficult problem combinatorially	✓	0.21
clustering is one of the steps in exploratory data analysis	✓	0.21
This paper presents an overview of pattern clustering methods from a statistical pattern recognition perspective	✓	0.33
We present a taxonomy of clustering techniques	✓	0.20
We also describe some important applications of clustering algorithms such as image segmentation, object recognition, an	✓	0.32

## References

- <https://doi.org/10.1038/sdata.2016.18>
- <https://doi.org/10.18653/v1/2022.acl-long.468>
- <https://doi.org/10.1145/331499.331504>