

# Explainable Multilingual Deep Text Clustering Using Transformer-Based Semantic Alignment

Sadhana Bobade<sup>1</sup>, Prof.Rupali Masake<sup>2</sup>

<sup>1,2</sup>Department of Computer Engineering Trinity College of Engineering And Research pune, S. NO. 25 27, Kondhwa-Saswad Road, Pisoli

---

## ABSTRACT

The increasing volume of multilingual textual data has created a need for efficient language-independent text clustering techniques. Traditional clustering approaches rely on surface-level features and cannot capture deep semantic relationships across languages. This survey presents a comprehensive review of explainable multilingual deep text clustering methods, focusing on transformer-based semantic alignment. This study examines recent advances in multilingual transformer models that enable shared semantic embedding spaces for cross-lingual text representation. This study analyzes various unsupervised clustering techniques applied to multilingual embeddings and highlights the importance of explainability in interpreting clustering results. Existing approaches for generating human-understandable explanations, such as key term extraction and semantic interpretation, are also reviewed. Furthermore, the survey identifies current challenges, including scalability, low-resource language support, and interpretability, and outlines potential future research directions for addressing these challenges. This survey aims to provide researchers with a clear understanding of state-of-the-art methods and emerging trends in explainable multilingual text clustering.

**Keywords:** multilingual text clustering, transformer models, semantic alignment, explainable AI, unsupervised learning, cross-lingual embeddings, transformer-based sentence encoding, layer normalization, K-means/hierarchical clustering, logistic regression for explanation, optional contrastive learning.

---

## INTRODUCTION

The exponential growth of digital content has led to the generation of massive amounts of textual data in multiple languages worldwide. Social media platforms, news portals, academic repositories, and online forums produce multilingual texts daily, making the efficient organization and analysis of such data a critical challenge. Text clustering plays a vital role by automatically grouping similar documents, thereby enabling applications such as topic discovery, information retrieval, and content recommendation.

Traditional text clustering techniques mainly depend on surface-level representations, such as bag-of-words, term frequency-inverse document frequency (TF-IDF), and n-grams. Although these methods are effective for monolingual datasets, they struggle to capture semantic meaning and contextual relationships, especially in multilingual environments. Language-specific features and vocabulary differences further limit their ability to perform accurate cross-lingual clustering tasks.

Recent advancements in deep learning, particularly in transformer-based models, have significantly improved text representation learning. Multilingual transformers, such as mBERT and XML-R, can generate deep contextual embeddings that encode semantic information across languages within a shared embedding space. These representations enable language-independent analysis and make cross-lingual text clustering more effective than previous methods. Despite these advances, most multilingual clustering systems function as black boxes and lack an explanation. The

The absence of human-understandable explanations reduces user trust and limits their applicability in sensitive domains such as healthcare, governance, and academic research. Explainable Artificial Intelligence (XAI) addresses this issue by providing

insights into the model decisions and clustering outcomes.

This survey paper explores explainable multilingual deep text clustering approaches based on transformer-based semantic alignments. It reviews existing methodologies, clustering techniques, and explainability strategies, identifies current challenges, and highlights future research directions. The goal of this survey is to provide a structured understanding of state-of-the-art developments, while emphasizing the importance of interpretability and language independence in multilingual text clustering systems.

## II. Related Work

Text clustering has been extensively studied in the field of Natural Language Processing (NLP), with early approaches relying on traditional machine learning techniques such as K-Means, Hierarchical Clustering, and Latent Dirichlet Allocation (LDA). These methods typically use surface-level text representations, including Bag-of-Words and TF-IDF. While effective for monolingual datasets, they fail to capture semantic and contextual information and perform poorly when applied to multilingual text because of vocabulary mismatch and language dependency.

To overcome these limitations, cross-lingual and multilingual text representation techniques have been introduced. Early multilingual approaches employed machine translation or bilingual dictionaries to map documents into a common language, prior to clustering. Although these methods improve cross-lingual performance, they are computationally expensive and prone to translation errors, especially for low-resource languages.

The emergence of word embedding models such as Word2Vec and GloVe enabled the semantic representation of words, and later extensions such as multilingual word embeddings allowed alignment across languages. However, these models provide static embeddings and cannot capture contextual meanings, limiting their effectiveness for complex textual data. Recent advancements in transformer-based models have significantly improved multilingual text processing. Pretrained models such as Multilingual BERT (mBERT), XLM, and XLM-RoBERTa generate deep contextual embeddings that encode semantic relationships across multiple languages in a shared embedding space. Several studies have demonstrated that these embeddings improve performance in multilingual text clustering and cross-lingual similarity tasks without requiring parallel data or translations.

Despite improved clustering accuracy, most existing multilingual clustering approaches lack an explanation. Clustering decisions are often opaque, making it difficult to understand why documents are grouped. To address this, explainable AI techniques, such as keyword extraction, attention visualization, and topic-based interpretation, have been explored. Some recent studies have integrated post-hoc explanation methods to identify representative terms or semantic features contributing to cluster formation.

However, the explainability of multilingual clustering remains an underexplored area. Existing solutions often focus on accuracy while neglecting interpretability, scalability, and support for low-resource languages (LRLs). This survey builds upon prior research by analyzing transformer-based multilingual clustering methods with a specific emphasis on explainability, semantic alignment, and practical applicability, thereby highlighting the need for transparent and trustworthy multilingual text clustering systems.

## III. Challenges and Motivation

One major challenge is achieving accurate semantic alignment across languages with diverse grammatical structures and vocabularies. Multilingual transformers may still produce inconsistent representations, particularly for low-resource languages. Another challenge is integrating explainability into deep clustering models without degrading their performance. Interpreting high-dimensional embeddings and attention mechanisms in multilingual settings is complex and computationally demanding. Additionally, scalability remains an issue because of the high resource requirements of transformer models. Ensuring fair and unbiased clustering outcomes across different languages further adds to the complexity of the task.

The rapid growth of multilingual digital content has created a strong need for automatic text clustering methods that can operate across languages while preserving the semantic meaning. Traditional clustering techniques often fail to capture deep contextual relationships and effectively align semantics between different languages. Transformer-based multilingual models offer powerful representations, but their black-box nature limits their trust and usability. This study was motivated by the need to combine accurate multilingual clustering with explainable mechanisms that allow users to understand why texts from different languages are grouped together. Providing transparent and interpretable clustering results is essential for reliable deployment in real-world and cross-cultural settings.

#### IV. Proposed Methodology

The proposed methodology presents an explainable and language-independent framework for multilingual text clustering using transformer-based semantic alignment. The system is designed to cluster multilingual documents in an unsupervised manner while providing human-understandable explanations for the generated clusters. The overall workflow of the proposed approach is as follows:

##### 1. Data Collection and Preprocessing

Multilingual textual data are collected from diverse sources, such as news articles, social media posts, and document repositories. Basic preprocessing steps were applied, including text normalization, noise removal, and handling of special characters. Unlike traditional approaches, language-specific preprocessing was minimized to maintain language independence.

##### 2. Multilingual Text Representation

Pretrained multilingual transformer models, such as Multilingual BERT (mBERT) or XLM-RoBERTa, are used to generate deep contextual embeddings for each document. These models map texts from different languages into a shared semantic embedding space, enabling effective cross-lingual alignment without the need for translation or labeled data.

##### 3. Semantic Alignment

The extracted embeddings inherently capture semantic similarities between languages. This shared representation ensures that documents discussing similar topics are placed close to each other in the embedding space regardless of the language used. Semantic alignment is the foundation of accurate multilingual clustering.

##### 4. Unsupervised Clustering

Clustering algorithms, such as K-Means or Hierarchical Clustering, are applied to the aligned multilingual embeddings. These algorithms group documents based on semantic similarity, producing meaningful clusters without requiring prior knowledge of the number of topics or labeled training data points.

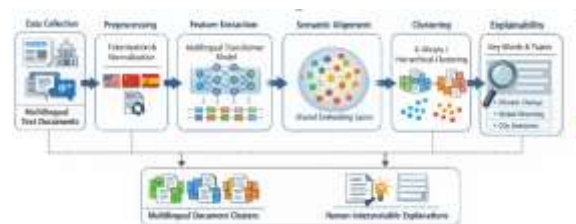
##### 5. Explainability Module

To enhance the interpretability, an explainability module was integrated into the framework. This module identifies the key indicative words, phrases, or semantic features that characterize each cluster. Techniques such as term importance scoring and embedding-based keyword extraction are used to generate human-understandable explanations for the formation of clusters.

##### 6. Evaluation and Analysis

The clustering performance was evaluated using standard metrics, such as the Silhouette Score, Davies–Bouldin Index, and qualitative analysis of cluster coherence. The quality of the explanations was assessed based on the interpretability and relevance of the extracted keywords. The proposed methodology ensures accurate multilingual text clustering while maintaining its transparency and explainability. By combining deep semantic understanding with interpretable insights, the framework addresses the key limitations of existing multilingual clustering systems and supports real-world applications involving diverse and low-resource languages.

#### V. Proposed System Architecture



#### VI. Algorithm Working

- Transformer-based Sentence Encoding
- Layer Normalization

- K-Means / Hierarchical Clustering
- Logistic Regression for Explanation
- Optional Contrastive Learning

#### Step 1: Input Acquisition

The system takes a collection of unlabeled multilingual text documents as its input. The documents may belong to different languages and domains.

#### Step 2: Text Preprocessing

Each input document underwent basic preprocessing, including text normalization, removal of unnecessary symbols, and noise reduction. Language-specific preprocessing was kept to a minimum to preserve language independence.

Step 3: Feature Extraction Using Multilingual Transformers Pretrained multilingual transformer models, such as mBERT or XLM-R, are used to convert each document into a dense contextual embedding. These embeddings capture deep semantic and contextual information across languages.

#### Step 4: Semantic Alignment

The generated embeddings were mapped to a shared semantic space, where documents with similar meanings were positioned closer together, irrespective of their original language.

#### Step 5: Unsupervised Clustering

An unsupervised clustering algorithm (such as K-Means or Hierarchical Clustering) is applied to the aligned embeddings to group semantically similar documents into clusters.

#### Step 6: Cluster Labeling and Analysis

Each cluster was analyzed to identify dominant semantic patterns. The system computes representative features for each cluster to understand its theme.

#### Step 7: Explainability Generation

An explainability module extracts key indicative words or phrases from each cluster using importance scoring or embedding-based keyword-extraction techniques. These keywords serve as human-understandable explanations for the formation of clusters.

#### Step 8: Output Generation

The final output consists of clustered multilingual documents with corresponding explanations in the form of representative keywords or topic summaries.

This algorithm ensures accurate multilingual text clustering while providing transparency and interpretability, making the system suitable for real-world and research-oriented applications in the future.

### VII. Proposed Working

The proposed system works by integrating multilingual transformer-based semantic understanding with unsupervised clustering and explainability techniques to achieve this goal. The overall operation of the system is described below.

Initially, the system accepts a collection of multilingual text documents as the input. These documents may belong to different languages and domains and do not require labeled information. The input text is then passed through a preprocessing stage, where basic cleaning and normalization are performed to remove noise while maintaining language independence.

After preprocessing, the cleaned text is fed into a pretrained multilingual transformer model, such as Multilingual BERT (mBERT) or XLM-RoBERTa. This model converts each document into a deep, contextual vector representation. Because the transformer is trained on multiple languages, it projects all documents into a shared semantic embedding space, enabling effective cross-lingual alignment. The aligned embeddings are then processed by an unsupervised clustering algorithm, such as K-Means or Hierarchical Clustering. Based on semantic similarity, the algorithm groups related documents into meaningful clusters without requiring prior knowledge of the document labels or languages.

An explainability module was integrated into the system to improve transparency and trust. This module analyzes clustered documents to extract key indicative words, phrases, or semantic features that characterize each cluster. These extracted terms provide human-understandable explanations that help users understand why documents are grouped.

Finally, the system produces clustered multilingual documents with corresponding explanations. The proposed framework ensures accurate semantic clustering, language independence, and interpretability, making it suitable for real-world multilingual text analysis and research applications.

### **CONCLUSION**

This paper presents an in-depth study of explainable multilingual deep text clustering with a focus on transformer-based semantic alignment. With the rapid increase in multilingual textual data, traditional clustering methods are no longer sufficient because of their inability to capture deep semantic and contextual relationships across languages. The reviewed approaches demonstrate that multilingual transformer models effectively generate shared semantic representations, enabling accurate and language-independent text clustering in an unsupervised way.

This proposal highlights the importance of integrating explainability into multilingual clustering systems to improve transparency, trust, and usability. By incorporating explainable AI techniques, such as keyword extraction and semantic interpretation, the clustering results become more understandable and meaningful to users. The analysis also identified key challenges, including scalability, interpretability in low-resource languages and computational complexity.

Overall, this study emphasizes that combining transformer-based multilingual embeddings with explainable clustering techniques offers a promising direction for organizing and analyzing large-scale multilingual text data. Future advancements in this area are expected to further enhance the clustering accuracy, interpretability, and applicability across diverse real-world domains.

### **ACKNOWLEDGMENT**

This work on Explainable Multilingual Deep Text Clustering Using Transformer-Based Semantic Alignment was supported by academic research facilities and institutional resources that enabled the successful completion of this study.

### **REFERENCES**

- [1.] J. Euzenat and P. Shvaiko, "Ontology Matching," in *Ontology Matching*, Springer, 20020.
- [2.] L. Ehrlinger and W. Woß, "Towards a definition of knowledge graphs," *SEMANTiCS*, 2016.
- [3.] I. Horrocks, B. Motik, and R. Shearer, "The Hermit OWL Reasoner," in *Proceedings of OWLED*, 2021.
- [4.] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1–2, pp. 83–97, 2024.
- [5.] E. Jimenez-Ruiz and B. Cuenca Grau, "LogMap: Logic-based and scalable ontology matching," in *Proceedings of the 10th International Semantic Web Conference (ISWC)*, 2011, pp. 273–288.
- [6.] D. Faria, C. Pesquita, E. Santos, M. Palmonari, I. F. Cruz, and F. M. Couto, "The AgreementMakerLight ontology matching system," in *On the Move to Meaningful Internet Systems: OTM 2013 Conferences*, Graz, Austria, 2013, pp. 527–544.
- [7.] M. Chen, Y. Tian, M. Yang, and C. Zaniolo, "Multilingual knowledge graph embeddings for cross-lingual knowledge alignment," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, Melbourne, Australia, 2017, pp. 1511–1517.
- [8.] Z. Wang, Q. Lv, X. Lan, and Y. Zhang, "Cross-lingual knowledge graph alignment via graph convolutional networks," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Brussels, Belgium, 2018, pp. 349–357.