International Journal of Research in Engineering, Science and Management
Volume 9, Issue 1, January 2026
www.ijresm.com | E-ISSN: 2581-5792 | RESAIM Publishing | www.resaim.com

Comparative Analysis of TF-IDF and LLM-Based Text Corpus Generation for Course Recommendation Systems

Deepanshu Aggarwal[1*], Neerja Doshi[1], Sanskar Kamble[1], Sudhir Dhage[1]
[1]*Department of Computer Engineering, Sardar Patel Institute of Technology, Mumbai, India*

***Abstract*: Course recommendation systems have become increasingly critical in addressing the complex educational challenges of personalized learning pathways. This research introduces a novel approach to text corpus generation that leverages advanced natural language processing techniques to enhance course recommendation accuracy. Traditional keyword extraction methods, particularly TF-IDF, often struggle to capture nuanced semantic relationships within academic content. Our study proposes an innovative methodology that combines traditional statistical methods for large language model (LLM) keyword extraction to generate a comprehensive, multidimensional course corpus. The proposed framework segments college databases across semesters and subjects, generating a three-dimensional keyword representation that captures the intricate relationships between academic content. By comparing traditional TF-IDF keyword extraction with LLM-based semantic keyword generation, we demonstrate significant improvements in recommendation relevance and precision. Experimental results reveal that LLM-based approaches substantially outperform traditional statistical methods in capturing contextual and semantic nuances of academic content.**

***Keywords*: Recommendation system, Keyword extraction, TF-IDF, nDCG Evaluation metric, Fine-Tuned LLM, Natural Language Processing (NLP).**

1. Introduction

Higher education’s digital transition has reached a turning point that calls for creative answers to the complex problems of navigating the new academic environment. Due to the long-standing fragmentation of traditional academic support systems, students must navigate a maze of disjointed platforms, which makes it difficult to make well-informed decisions. This study addresses a basic issue in the educational system: the crucial but sometimes disregarded course selection process. Selecting a course is a calculated decision that significantly affects a student’s academic career and future career prospects. It is more than just an administrative task. However, the modern world is characterized by a lack of personalization and opacity of information. Less than ideal academic experiences and probable misalignment between individual potential and chosen courses might arise from students’ typical reliance on arbitrary choices, peer recommendations, or limited contextual understanding.

Our research introduces a transformative approach to course recommendation using advanced artificial intelligence and natural language processing technologies. By developing a comprehensive, integrated student dashboard, we aim to revolutionize how students interact with their academic opportunities. The cornerstone of this innovation is a sophisticated Large Language Model (LLM)-powered recommendation system that transcends traditional recommendation methodologies. The proposed system distinguishes itself through a novel text corpus generation methodology. We present a dual-approach strategy for keyword extraction, comparing traditional statistical methods like term frequency-inverse document frequency (TF-IDF) with advanced LLM-based semantic analysis. This approach ensures a robust and nuanced understanding of course content, enabling highly personalized recommendations that align with individual student profiles.

By bridging technological innovation with educational strategy, this research addresses a critical gap in current academic support systems. We demonstrate how advanced machine learning techniques can transform course selection from a challenging, uncertain process to a data-driven, personalized journey of academic exploration.

2. Related Work

In recent years, large-language models (LLMs) have shown significant promise in recommendation systems. In [1], tekim 2024 large an LLM-based recommender system is introduced to address the limitations of traditional collaborative filtering, especially in cold start scenarios where collaborative filtering typically fails. This work is relevant as it highlights the need for more robust models in recommendation tasks.

Early studies, such as those in [2] and [3], focus on collaborative filtering-based course recommendation systems. These methods, while effective in some cases, struggle with sparse data and dynamic content, which is a limitation that the proposed LLM-based models aim to overcome.

The potential of LLMs in recommendation systems has also been explored in [4], where they discuss the use of OpenAI-

*Corresponding author: deepanshuaggarwal023@gmail.com

GPT with in-context learning, adapting to new tasks or information based on the input prompt. Additionally, in [1], tebao 2023 tallrec propose a framework, TALLRec, for fine-tuning LLMs using recommendation data, showing improved performance in cold-start and cross-domain scenarios. This aligns well with our approach to fine-tuning LLMs for course recommendations.

In the area of explainable recommendations, in [6] pro- pose Chat-rec, an interactive and explainable LLM-augmented recommender system. This work is notable for its focus on increasing the interpretability of recommendations, which is crucial in educational settings where students need to understand the reasoning behind course suggestions.

In [7] explore the use of LLMs for generating item descriptions in recommendation systems, eliminating the need for extensive web scraping. This is similar to our approach, where LLMs generate content-based recommendations, such as course descriptions, by understanding the course syllabus. Keyword extraction plays a critical role in generating relevant recommendations in [8] comparing TF-IDF and log- likelihood methods for keyword extraction, demonstrating that TF-IDF outperforms log-likelihood in accuracy for keyword extraction. This supports our use of TF-IDF in the initial stages of corpus creation for course recommendations. Contributions from [13] highlight that TF-IDF can be leveraged to uncover emerging research themes in Indonesian grant funding, high- lighting the potential of this technique for identifying trends in academic research.

On the efficiency of fine-tuning LLMs for recommendation tasks, such as the work by Lin et al. in [9] and in [10], shows how selecting influential data and using LLM-generated responses for fine-tuning can significantly enhance performance. This paper also suggests that fine-tuned LLMs work best when they are inputted with responses that are LLM-generated.

When talking about evaluation metrics in recommendations especially when the ranking of recommendation is most required; [11] discusses that nDCG is a good metric. It also opens the discussion of *nDCG@10* which takes the ranking of 10 recommendations. Adding to this contributions from [12] are useful as well.

Corpus Generation for LLM requires *Data to Text* i.e., extracting data from Tables, Graphs and making it in machine-understandable format. [16] suggests methods to achieve it in great detail. To scale LLMs efficiently [17] suggests corpus sizes and how to extract influential data to feed to LLM.

A. Type of LLM

In this research, we explore different approaches for using language models (LLMs) for the specific task of course recommendation, which is primarily centred around extracting relevant courses based on extracted keywords from academic syllabi. In [15], we come across various types of LLMs. The three main types of LLM which can be highlighted are :

- *Pre-trained LLM:* A *pre-trained language model* has been trained on vast datasets containing general knowledge from diverse sources such as books, articles, web- sites, and other publicly available text corpora. These models are powerful at understanding and processing natural language, but they are not tailored for specific domains like course recommendations.
  Example: OpenAI's GPT series (such as GPT-3 or GPT- 4) and Google's BERT.
- *Fine-tuned LLM: Fine-tuning* a pre-trained language model involves further training it on a specific domain or dataset to make the model more suitable for particular tasks. In our case, we employ fine-tuning by adapting a general-purpose pre-trained LLM to a specialized course recommendation system. [7] This domain-specific dataset allows the model to better understand the relationships between course content, prerequisites, and subject terminology, and how they relate to each other, ensuring that course recommendations are both contextually relevant and accurate.
  Example: Fine-tuning OpenAI's GPT or BERT with academic syllabi, course descriptions, and keyword data.
- *Few-shot Learning: Few-shot learning* is a method where the model is trained on a minimal amount of task-specific data. [14] Instead of extensive training, the model learns to perform the task by being provided with a few examples or prompts that highlight the structure and intent of the task.
  Example: A prompt might contain examples of how to recommend courses based on extracted keywords (e.g., "Recommend a course based on these keywords: Algorithms, data structures, machine learning").

In this study, fine-tuning a pre-trained LLM was chosen as the optimal approach for course recommendation.

3. Methodology

The methodology section outlines the comprehensive approach used in developing an advanced course recommendation system. This section presents a detailed exploration of two important keyword extraction methodologies: the traditional TF-IDF statistical approach and an innovative Large Language Model (LLM)-based semantic extraction technique. By comparing these methods to improve the accuracy of course recommendations, we want to shed light on the subtle variations in text corpus development. Our proposed methodology encompasses four critical components:

- Systematic syllabus data preprocessing
- TF-IDF statistical keyword extraction
- LLM-based semantic keyword generation
- Personalized recommendation ranking

A. Syllabus Data Preprocessing

The syllabus data is preprocessed using a structured format:

Semester X Subject 1
[Detailed Syllabus Content]
Subject 2

[Detailed Syllabus Content]

The parsing algorithm ensures:

- Semantic separation of semesters
- Clear subject demarcation
- Contextual information preservation

B. *Keyword Extraction Methodologies*

Two primary approaches were used:

- TF-IDF Statistical Extraction
- Large Language Model (LLM) Semantic Extraction

1) *TF-IDF Keyword Extraction*

The Term Frequency- Inverse Document Frequency (TF-IDF) method quantifies word importance through frequency and uniqueness. Mathematical Formulation:

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t) \qquad (1)$$

Where:

$$\text{TF}(t, d) = \frac{\text{Term Frequency}}{\text{Total Terms in Document}}$$

$$\text{IDF(t)} = \log \frac{\text{Total Documents}}{\text{Documents with Term}}$$

Algorithm 1 TF-IDF Keyword Extraction

```
procedure ExtractKeywords($corpus, k$)
    $result$ ← Initialize 3D Array
    for $semester$ ∈ $corpus$ do
        for $subject$ ∈ $semester$ do
            $scores$ ← Calculate TF-IDF($subject$)
            $top_k$ ← Select Top $k$ Keywords
            $result[semester][subject]$ ← $top_k$
        end for
    end for
    return $result$
end procedure
```

2) *LLM-Based Semantic Keyword Extraction*

Unlike tradi- tional statistical methods, the Large Language Model (LLM) approach leverages advanced semantic understanding and contextual intelligence for keyword extraction. This method addresses the limitations of frequency-based approaches by capturing deeper semantic relationships within the syllabus content.

a) *Chunking Strategy:* The syllabus is segmented into hierarchical chunks to facilitate efficient and contextually aware processing:

- *Primary Segmentation*: Divide syllabus by semesters
- *Secondary Segmentation*: Break each semester into subject-specific chunks
- *Tertiary Segmentation*: Further divide subjects into logical content units

Chunking Example:

Semester 3 [Primary Level]

Computer Networks [Secondary Level]

- Network Fundamentals
- Protocol Architectures
- Security Mechanisms [Tertiary Level]

b) *LLM Keyword Extraction Algorithm*

Algorithm 2 LLM Semantic Keyword Extraction

```
procedure LLMKeywordExtraction($corpus$)
    $semantic\ keywords$ ← Initialize 3D Array
    for each $semester$ ∈ $corpus$ do
        for each $subject$ ∈ $semester$ do
            $chunk\ content$ ← Preprocess($subject$)
            $prompt$ ← ConstructPrompt($chunk_{c}ontent$)
            $llm\ keywords$ ← QueryLLM($prompt$)
            $filtered\ keywords$ ← FilterRelevantKeywords($llm\ keywords$)
            $semantic\ keywords[semester][subject]$ ← $filtered\ keywords$
        end for
    end for
    return semantickeywords
end procedure
```

c) *Prompt Engineering:* The LLM extraction relies on carefully crafted prompts:

Analyze the following syllabus content and extract the top 50 most significant keywords. Focus on:

- Core learning objectives
- Essential conceptual domains
- Potential research or application areas

Keyword Scoring Mechanism:

$$S_k = \omega_1 \cdot \text{Semantic Relevance} + \omega_2 \cdot \text{Contextual Significance} \qquad (2)$$

Where:

S_k = Keyword Score

ω_1, ω_2 = Weighted Parameters

C. *Comparative Advantages*

The LLM approach offers several key improvements:

- Contextual understanding beyond mere frequency
- Capture of nuanced semantic relationships
- Adaptive to diverse academic domains
- Reduced dependency on predefined statistical metrics

D. *Recommendation Scoring*

We now evaluate the performance of our recommendation system using the averaged NDCG metric. The goal of the NDCG score is to give more weight to relevant recommendations appearing higher in the ranked list.

- *NDCG Formula*

The Normalized Discounted Cumulative Gain (NDCG) is used to evaluate the ranking quality of the recommendations. The formula for NDCG at rank k is given by:

$$NDCG@k = \frac{DCG@k}{IDCG@k}$$

Where:

$$DCG_k = \sum_{i=1}^{k} \frac{rel(i)}{\log_2(i+1)}$$

and *rel*(*i*) is the relevance score of the *i*-th recommendation, and $IDCG_k$ is the ideal DCG, calculated assuming the perfect relevance order.

The average NDCG score is calculated by considering the 10 recommendations made by the system with the user prompt and evaluating their relevance in terms of ranking to the user's query. We compute the NDCG score for two different text corpora: one generated using the TF-IDF technique and the other generated using the LLM-based approach.

4. Result and Analysis

An important step in comprehending the revolutionary potential of large language models (LLMs) in educational technology is the actual testing of our course recommendation system. The performance of two different keyword extraction methodologies—the conventional Term Frequency- Inverse Document Frequency (TF-IDF) approach and our suggested LLM-based semantic extraction technique—is carefully compared in this section's thorough analysis of the test data.

A. Performance Comparison of Keyword Extraction Methods

The comparative analysis between traditional TF-IDF and Large Language Model (LLM)-based keyword extraction revealed significant differences in course recommendation accuracy.

Table 1
Performance comparison of keyword extraction methods

Approach	NDCG@10	Improvement
TF-IDF	0.612	-
LLM-based	0.879	43.6%

1. *Quantitative Performance Metrics:* We evaluated the performance using the Normalized Discounted Cumulative Gain (NDCG) metric at rank 10:
2. *Statistical Significance:* A paired t-test confirmed the statistical significance of the performance difference (p-value<0.001), validating the substantial improvement offered by the LLM-based approach.

B. Key Findings

The LLM-based method demonstrated superior performance through:

- Enhanced contextual relevance
- Improved keyword diversity
- Greater adaptability across academic domains

5. Conclusion

A revolutionary development in course recommendation systems is the appearance of large language models (LLMs), offering unprecedented potential to personalize and enhance academic pathway selection. Our comparative study between traditional TF-IDF and LLM-based semantic keyword extraction methodologies reveals significant advancements in recommendation accuracy and contextual understanding.

The key findings of our research demonstrate the substantial superiority of the LLM-based approach:

- The LLM-based keyword extraction method achieved a remarkable 43.6% improvement in NDCG@10 metric compared to the traditional TF-IDF approach, with statistical significance confirmed by a paired t-test (p-value < 0.001).
- LLM-based techniques demonstrated superior performance through enhanced contextual relevance, improved keyword diversity, and greater adaptability across diverse academic domains.
- The proposed methodology successfully bridges techno- logical innovation with educational strategy, transforming course selection from an uncertain process to a data- driven, personalized academic exploration.

By using advanced artificial intelligence and natural language processing technologies, this research contributes to a more intelligent, adaptive, and student-centric approach to academic course recommendation systems.

References

[1] S. Kim *et al.*, "Large language models meet collaborative filtering: An efficient all-round LLM-based recommender system," in *Proc. 30th ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD)*, 2024.

[2] N. B. Salehudin *et al.*, "A proposed course recommender model based on collaborative filtering for course registration," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 11, 2019.

[3] K. K. Jena *et al.*, "E-learning course recommender system using collaborative filtering models," *Electronics*, vol. 12, no. 1, Art. no. 157, 2022.

[4] Z. Zhao *et al.*, "Recommender systems in the era of large language models (LLMs)," *arXiv preprint*, arXiv:2307.02046, 2023.

[5] K. Bao *et al.*, "TallRec: An effective and efficient tuning framework to align large language models with recommendation," in *Proc. 17th ACM Conf. Recommender Systems (RecSys)*, 2023.

[6] Y. Gao *et al.*, "Chat-Rec: Towards interactive and explainable LLMs-augmented recommender systems," *arXiv preprint*, arXiv:2303.14524, 2023.

[7] A. Acharya, B. Singh, and N. Onoe, "LLM-based generation of item descriptions for recommendation systems," in *Proc. 17th ACM Conf. Recommender Systems (RecSys)*, 2023.

[8] M. A. Abid *et al.*, "Comparative analysis of TF-IDF and log-likelihood methods for keyword extraction of Twitter data," *Mehran University Research Journal of Engineering and Technology*, vol. 42, no. 1, pp. 88–94, 2023.

[9] X. Lin et al., "Data-efficient fine-tuning for LLM-based recommendation," in Proc. 47th Int. ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2024.

[10] X. Ren, B. Wu, and L. Liu, "I learn better if you speak my language: Understanding the superior performance of fine-tuning large language models with LLM-generated responses," in *Proc. 2024 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2024.

[11] D. Cheng *et al.*, "The ENGAGE corpus: A social media dataset for text-based recommender systems," in *Proc. 13th Language Resources and Evaluation Conf. (LREC)*, 2022.

[12] Y.-D. Seo *et al.*, "Group recommender system based on genre preference focusing on reducing the clustering cost," *Expert Systems with Applications*, vol. 183, Art. no. 115396, 2021.

[13] R. Setiawan, Z. Kisman, and A. Imam, "Analysis of frequently appearing words in the titles of 2023 research grant winners in Indonesia using the TF-IDF method," *Journal of Information Systems and Informatics*, vol. 5, no. 4, pp. 1508–1522, 2023.

[14] T. B. Brown *et al.*, "Language models are few-shot learners," *arXiv preprint*, arXiv:2005.14165, 2020.

[15] Z. Guo *et al*., "Evaluating large language models: A comprehensive survey," *arXiv preprint*, arXiv:2310.19736, 2023.

[16] J. Mahapatra and U. Garain, "Impact of model size on fine-tuned LLM performance in data-to-text generation: A state-of-the-art investigation," *arXiv preprint*, arXiv:2407.14088, 2024.

[17] B. N. Kaushikk, "Scaling efficient large language models," *arXiv preprint*, arXiv:2402.14746, 2024.