applied sciences

MDPI

Article

Mapping Moodle Resources to Course Topics Using Text Similarity Methods and Expert Evaluation

Beata Gancevska * and Simona Ramanauskaitė

Department of Information Technologies, Faculty of Fundamental Sciences, Vilnius Gediminas Technical University, 10223 Vilnius, Lithuania; simona.ramanauskaite@vilniustech.lt
* Correspondence: beata.gancevska@vilniustech.lt

Abstract

In this research, the alignment and mapping between Modular Object-Oriented Dynamic Learning Environment (Moodle) learning resources and course topics are described using text similarity methods. The goal of this work is to improve the accuracy of automated alignment between Moodle course learning resources and course topics by analyzing text similarity method performance and examining factors that affect how closely they match expert evaluation. During this research, an expert first mapped the e-course learning resources to course topics, after which multiple text similarity techniques were applied to match resource titles and descriptions to those topics. The findings show that the Large Language Model (LLM)-based solution achieves the lowest mean absolute error (MAE), the lowest mean squared error (MSE), and the strongest agreement with expert evaluation. Traditional keyword-based methods, such as Jaccard similarity and Term Frequency–Inverse Document Frequency (TF-IDF), demonstrate moderate performance, while the Sentence Bidirectional Encoder Representations from Transformers (SBERT)-based model shows the weakest alignment with expert evaluation.

Keywords: Moodle; text similarity; course resource-topic alignment; natural language processing; expert evaluation

1. Introduction

Learning in the virtual environment has become an important part of education, and one of the most widely used learning management systems (LMS) in the world is Moodle (Moodle Pty Ltd., Perth, Australia) [1–3]. In this study, the Moodle LMS was selected as the research platform due to its wide usage in higher education, particularly in universities, its open-source nature, and its flexibility in supporting different teaching, assessment, and learning types. Secondly, Moodle provides instructors with the possibility to create learning resources characterized by a resource name and an optional textual description, to which text similarity methods can be applied. This structure is compatible with the representation of course topics in official curriculum documentation, such as the study subject card, where course content is defined through a set of predefined topics.

In practice, especially in large-scale or long-running courses, instructors frequently edit learning resources and topics and add new materials incrementally during course delivery. For this reason, learning resources may not always remain aligned with the predefined course topics specified in the course syllabus. In such situations, misalignments may occur between the content of Moodle learning resources and the planned course topics in the curriculum. Consequently, Moodle-based courses need to be analyzed, as

Check for updates

Received: 10 January 2026
Revised: 14 February 2026
Accepted: 17 February 2026
Published: 19 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

these misalignments can negatively affect course coherence and the quality of learning, highlighting the need to evaluate resource–topic alignment in higher education.

Another aspect that might indicate the need for automated matching between Moodle courses and the course syllabus is online course delivery. While in-person taught courses always have a tutor who explains the Moodle course structure and resource presentation, fully online courses must be self-explanatory and reflect the course content even without viewing the resources. To ensure that the Moodle course matches the course content and forms a realistic impression of the course topics, it is important to estimate its alignment with the syllabus.

Digital learning platforms are used to provide learning materials for students, create assessments, enable communication and other activities, and therefore the coherence and structure of online courses are a critical part of course design [4,5]. One aspect of a well-structured course can be defined as a course where there is clear alignment between learning resources and the course topics. When a course matches this aspect, it means that all planned topics are present in the e-course and students can gain knowledge and achieve the learning outcomes. This evaluation is usually done manually, it is human-based, and as a result it is time-consuming, requires human resources, and depends on the subjective opinion of the person who performs the matching.

There are previous studies in which various aspects of learning resource organization and semantic similarity in educational contexts were analyzed. When doing research in learning analytics and educational data mining, natural language processing (NLP) techniques were used to classify learning materials, extract topics or recommend learning resources [6,7]. Text similarity techniques such as TF-IDF, cosine similarity, word embeddings and transformer-based models have been used to match textual content to predefined categories [8–10]. There are also studies where curriculum or syllabus mapping is performed [11,12], and others propose automated methods for detecting inconsistencies, redundancies, or duplicate content in learning materials using machine learning and text similarity techniques [13].

Overall, existing technical studies show that text similarity and machine learning methods have strong potential for analyzing educational content. Existing research has analyzed lexical similarity approaches, semantic embeddings, and transformer-based models for text classification and topic extraction in educational contexts [6–10]. However, most existing works focus on automated classification or recommendation tasks, while curriculum analysis is typically performed without incorporating expert judgment. In addition, there is a lack of studies that jointly evaluate lexical similarity methods, semantic embedding models, and large language model-based approaches within a single experimental framework and compare their results with expert mappings in real-world Moodle courses.

Existing studies do not compare Artificial Intelligence (AI) generated mappings of Moodle learning resources to course topics with mappings created by human experts. In addition, there is a lack of studies analyzing how different text similarity methods align with expert judgments in Moodle resource–topic mapping tasks. This research addresses this gap by analyzing the use of text similarity methods for aligning Moodle learning resources with course topics and comparing the resulting mappings with expert judgments.

The main aim of this study is to investigate how text similarity methods can support the automated assessment of alignment between Moodle learning resources and course topics. This study is important because it contributes to the automated assessment of course structure and consistency in LMS, helping course creators and evaluators identify redundancies and misalignments between learning resources and course topics, while also supporting course review and improvement.

2. Materials and Methods

2.1. Research Design

This research was conducted to evaluate how different text similarity methods can match Moodle learning resources with course topics from the course syllabus, and it is designed as a quantitative comparative analysis. In this study, human expert judgments are used as a reference for evaluating automated text similarity methods applied to align Moodle learning resources with course topics. Multiple text similarity methods are analyzed to identify which approaches produce results most consistent with expert evaluations. This approach is suitable for analyzing alignment accuracy because it compares multiple algorithms with expert opinion.

The process of this research consisted of several stages, which are shown in Figure 1. First, data from four Moodle LMS e-courses were collected, including all learning resource titles and descriptions, as well as the list of official course topics from the syllabus. Second, an expert created a resource–topic alignment matrix by assigning a percentage of alignment between each learning resource and each course topic. This expert-created matrix was used as a structured expert benchmark for evaluating the automated methods.

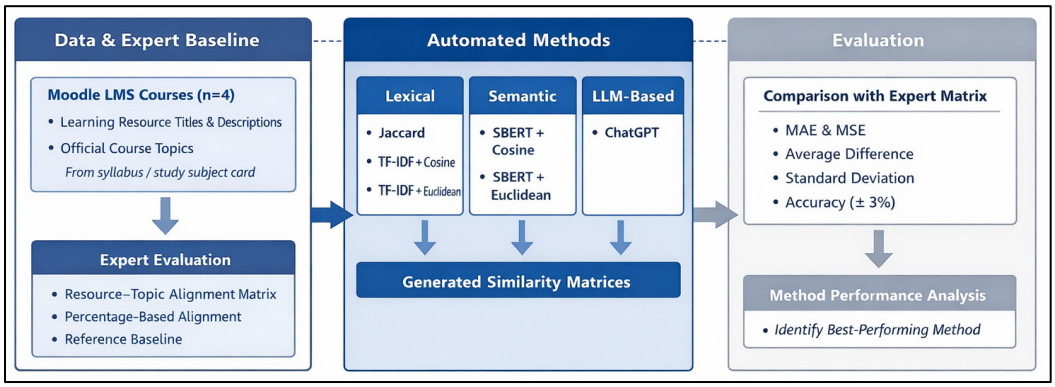


Figure 1. Overview of the research process.

In the next stage, multiple text similarity methods from three categories were applied: lexical (Jaccard similarity, TF-IDF with cosine similarity, TF-IDF with Euclidean distance), semantic (SBERT with cosine similarity, SBERT with Euclidean distance), and LLM-based (ChatGPT (OpenAI, San Francisco, CA, USA)). The goal was to automatically calculate similarity scores between course resources and course topics and, as a result, generate a similarity matrix comparable to the manually created expert evaluation matrix. Using methods from different categories allows for a more detailed and comprehensive comparison.

In the last stage, the matrices produced by each text similarity method were compared with the expert matrix using several metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), the average difference, the standard deviation of differences, and an accuracy measure showing how many method-generated values differed from the expert results by no more than three percentage points. This approach allows the identification of the method closest to human evaluation. Such an evaluation strategy is commonly used in text similarity research, where algorithmic outputs are compared against human judgement.

2.2. Dataset Description

The data for this research were collected from four higher education Moodle LMS e-courses. To reflect course variation, two computer science study program course owners were invited to collaborate. However, participation in this experiment was low. This

limitation was caused by course owners' lack of awareness or unwillingness to collaborate. In most cases, concern about being evaluated was observed, and course owners indicated that the course would be shared after planned improvements had been made.

For each course, learning resource titles and descriptions were automatically extracted from the Moodle database using SQL queries, while official course topics were obtained from the university information system. Learning resource titles were often very short and generic, such as "Final exam", "Laboratory Work No. 1", or "Homework". Resource descriptions frequently contained limited semantic information or technical content, including Hypertext Markup Language (HTML) fragments (e.g., "General news and announcements", "<hr/>", " You have 60 min "). In many cases, the combined resource text consisted of a generic title paired with a minimally informative description, for example the title "Final exam" with the description " You have 60 min ".

To conduct the study, separate Excel files were prepared for each course. The structure of these files was the same: the rows represented the learning resources with their titles and descriptions, and the columns represented the course topics. The structure of the Excel files is shown in Figure 2. At this stage, the matrices were empty and served as templates for the later stages of the research.

Resource Title	Resource Description	Course Topics	
		Installation of Hadoop and related technologies.	Python language and its syntax.
Midterm exam no 1 (2022-03-08)	Just Python basics: - variables - if - for - lists - dictionaries - arithmetic operations - work with files 		
Network file systems			

Figure 2. Structure of the Excel matrix for mapping learning resources to course topics.

These matrices provided a consistent format for comparing the expert results with the outputs of the text similarity methods. Once the expert filled in the alignment values, this structure allowed all results to be analyzed using different evaluation metrics.

Each generated matrix represents a single course. Therefore, four courses result in four separate matrices. The number of learning resources and course topics varies across courses, reflecting differences in course size and structure. Some courses contain a relatively small number of learning resources, while others include a larger and more diverse set of materials. Resource texts also differ in length and level of detail: some resources consist only of short or generic titles, whereas others include additional descriptive information. This variation reflects realistic Moodle course designs, where some courses provide limited material with minimal metadata, while others contain a larger number of resources with richer descriptions. The main characteristics of the dataset are summarized in Table 1.

Table 1. Main characteristics of the dataset.

Course ID	Number of Resources	Number of Topics	Resources with Description
807	60	19	25
2858	22	13	8
3941	155	20	30
6821	21	15	7

These Moodle courses were intentionally selected to represent realistic course configurations commonly used in higher education. The selected courses differ in the number of learning resources, the number of topics, and the availability of resource descriptions, which allows the evaluation of text similarity methods under different levels of information detail and course structural complexity. Including both smaller and larger courses, as well as courses with limited and more detailed resource descriptions, makes it possible to identify the strengths and limitations of different automated text similarity methods under diverse practical conditions.

In practice, Moodle courses are often updated during course delivery, and learning resources are added or modified incrementally. As a result, learning resources differ significantly in the amount and quality of available textual information. This selection was therefore motivated by the aim of analyzing how different text similarity methods perform in realistic Moodle course settings used in practice, rather than relying on standardized courses.

All data used in this research were fully anonymized. No student information, personal data, grades, or any other sensitive content were included. Therefore, the dataset is suitable for further analysis.

2.3. Expert Mapping Procedure

The expert's task was to evaluate the alignment between Moodle learning resources and course topics based on the resource titles, descriptions and the list of topics from four courses. For this purpose, the prepared Excel matrices were used as templates. Each matrix represented one course and was initially empty, containing only the structure needed for entering the alignment values.

In the mapping stage, the expert analyzed every learning resource title and description together with the list of course topics. In this study, the expert was a study program administrator with comprehensive knowledge of course structures, study plans, and quality assurance processes. In total, 2 experts were included in the research, each providing data independently for the courses included in the managed study program. Course instructors were not selected as experts, as they may be inclined to evaluate their own courses more positively, for example by assuming that all topics are sufficiently covered, no redundancies exist, or no misalignments are present. In contrast, the selected experts have a broader, program-level perspective on course structure requirements and quality standards, allowing for a more objective assessment of alignment between learning resources and course topics.

To evaluate how strongly each resource was related to each topic, the expert assigned a value from 0 to 100 percent for every resource-topic pair. The higher the value, the stronger the connection between the learning resource and the course topic, and vice versa. As a result, a resource–topic matrix was created for each course, where every cell contained the expert–assigned percentage value. An example of an expert matrix is presented in Figure 3.

Resource Title	Resource Description	Course Topics	
		Installation of Hadoop and related technologies.	Python language and its syntax.
Midterm exam no 1 (2022-03-08)	Just Python basics: - variables - if - for - lists - dictionaries - arithmetic operations - work with files 	0%	90%
Network file systems		50%	0%

Figure 3. Expert resource–topic matrix.

The analysis of expert mapping results shows that a substantial proportion of learning resources could not be aligned with any course topic. Across the analyzed courses, between 46.7% and 66.7% of resources received zero scores for all topics, indicating that these resources were too abstract or insufficiently described to allow reliable topic assignment. In contrast, only 33.3% to 53.3% of learning resources were aligned with one or more course topics. The detailed results are presented in Table 2.

Table 2. Expert mapping characteristics.

Course ID	Number of Resources	No Alignment (%)	With Alignment (%)
807	60	46.7	53.3
2858	22	54.5	45.5
3941	155	65.2	34.8
6821	21	66.7	33.3

A high proportion of zero scores was observed because the analyzed courses contain learning resources with administrative or assessment-oriented purposes, such as announcements, exams, homework assignments, or activities labeled in a generic manner (e.g., "Theory") without descriptive metadata. In real-world Moodle courses, many such resources are not directly associated with official course topics defined in the syllabus, as they primarily serve organizational or assessment functions rather than delivering specific learning content.

This demonstrates a lack of university and course owner orientation toward fully online courses, where each resource should be self-explanatory based on its title rather than requiring users to examine the full content. Currently, university studies are oriented toward live participation in classes, while Moodle is used as an additional platform for material sharing. However, when considering fully online course delivery, the obtained results are concerning and highlight the need for changes in course preparation methodology and requirements.

Resources with zero alignment scores were intentionally included in the analysis. Excluding these resources would simplify the dataset and could introduce bias by focusing only on clearly structured and well-documented learning materials. By preserving all learning resources, the analysis reflects realistic Moodle course conditions and enables the assessment of how text similarity methods handle non-topic-specific resources.

During the expert work, no automated methods or predefined rules were used, the evaluation was based only on human judgement and interpretation. These matrices serve as the reference benchmark for assessing the results produced by the automated text similarity methods.

2.4. Text Similarity Methods

In this research, several text similarity methods from different NLP categories were used to evaluate the alignment between Moodle learning resources and course topics. These included lexical similarity methods, semantic similarity methods and an LLM-based approach.

Lexical methods perform well with short texts [14,15], while semantic embedding methods capture deeper textual meaning. The LLM-based method was included to examine whether it can approximate human judgement [16,17]. Using all three categories makes it possible to identify the strengths and weaknesses of each approach and determine which method performs better for this type of data.

2.4.1. Lexical Similarity Methods

Lexical text similarity methods are based on word overlaps and term frequency [18,19]. Three methods from this category were applied in this study: Jaccard similarity [20,21], TF-IDF with cosine similarity [22,23], and TF-IDF with Euclidean distance [24].

Lexical similarity methods were implemented using the scikit-learn library, which was applied for text vector representation and similarity computation. To identify the most accurate configuration, different preprocessing approaches (e.g., stop words, n-grams) were tested, and the highest-performing sets were selected for further comparison. These methods are suitable for the short, context-limited texts characteristic of the Moodle dataset.

2.4.2. Semantic Similarity Methods

Methods from this category analyze text meaning rather than word overlap [25]. In this study, SBERT, a transformer-based model that represents text as vectors [26,27], was used to generate embeddings for learning resource and course topic texts. Two distance metrics were applied: cosine similarity and Euclidean distance.

Embeddings were generated using the sentence-transformers library. Several models were tested, and paraphrase-multilingual-MiniLM-L12-v2 was selected based on its accuracy. Its multilingual capability was particularly suitable, as some resource descriptions contained both English and Lithuanian text. The impact of model size was not significant, as the course language is domain-specific and the context is presented similarly across models. Testing a course-topic-related BERT model was not possible due to limited availability and multilingual constraints.

Semantic similarity methods are suitable for texts with richer context and more detailed information. Applying these methods to short texts with limited context, such as the dataset used in this study, resulted in weaker performance.

2.4.3. LLM-Based Method

A large language model-based solution, ChatGPT (version 5.2), was also used to generate resource–topic matrix values. The approach was selected due to the model's reasoning capabilities [28,29].

For the LLM-based method, the prompt for ChatGPT was formulated using the R.A.C.E. (Role, Action, Context, Expectation) framework. This framework was used to clearly define the task, scope, expected results, and output format, and ensure consistent model behavior across courses. The prompt was structured as follows:

- Role: You are an educational technology specialist with experience in Moodle content analysis.
- Action: Evaluate the alignment between Moodle learning resource titles and descriptions and course syllabus topics using a large language model.
- Context: You are provided with an Excel file containing Moodle learning resource titles and descriptions, as well as the corresponding course syllabus topics.
- Expectation: Generate a percentage-based alignment matrix in which each value indicates how strongly a Moodle learning resource aligns with a specific course topic. Provide the results in an Excel file.

For this task, the model was provided with an empty Excel matrix containing learning resources and course topics. As a result, it generated a matrix with filled values for each resource–topic pair, in a way like human evaluation. Using this technique, the study tested whether a reasoning-based model can process texts with limited semantic information.

The ChatGPT web interface was used to perform the evaluation. Direct control over generation parameters such as temperature, top-p sampling, or random seed was not analyzed, as it is usually related to answer generation quality rather than the provided data analysis. Across all analyzed courses the same prompt formulated using the R.A.C.E. framework was applied. Prompt engineering was conducted to find the best matching prompt for obtaining a structured response for the task.

This approach was used to reduce output variability and improve procedural consistency by employing a fixed and clearly structured prompt. The same prompt formulation was consistently used for all courses. However, because the web interface does not allow full control over generation parameters, complete reproducibility cannot be guaranteed. The described setup provides a consistent evaluation procedure within the analyzed dataset, but minor variability across runs remains possible.

2.5. Mapping Generation

The mapping between learning resources and course topics was created by applying the text similarity methods to the prepared dataset for each Moodle course. First, the text of each learning resource was preprocessed: the resource title and description were merged into a single text field. This was done because resource titles are often short or very generic (e.g., “Laboratory Work No. 1”) and do not clearly indicate the corresponding course topic. Combining titles with resource descriptions provides additional contextual information, resulting in a more informative representation for alignment analysis. After that, the text was cleaned, including the removal of HTML tags and other unnecessary elements.

In the next step, each learning resource was compared with each course topic, and similarity scores were produced by the selected methods. This process followed a many-to-many approach, where every resource was evaluated against every topic.

In this stage, a matrix was produced for each course by each text similarity method. Every matrix had the same structure consisting of rows, columns and cells:

- Learning resources were represented by the rows;
- Course topics were represented by the columns;
- Each cell contained the similarity score produced by the specific method.

The generated matrices could be evaluated together with the expert-created matrices because all of them shared the same structure and scale. For each course, four matrices were produced using different text similarity methods. These matrices are required for the next step, where each method is evaluated and compared with the expert results.

2.6. Evaluation Metrics

Several quantitative metrics were used to evaluate how closely each text similarity method matched the expert evaluation. Using these metrics, the expert matrix was compared with the matrices generated by the automated methods.

The selected metrics were: mean absolute error (MAE), mean squared error (MSE), average difference, standard deviation of differences, and accuracy. MAE and MSE quantify the magnitude of deviations between expert and automated scores, while the average difference and its standard deviation indicate systematic bias and variability across resource–topic pairs. Accuracy reflects the proportion of cases where the normalized score difference did not exceed three units.

When used together, these metrics provide a more detailed view of how closely automated methods align with expert judgments. Each metric captures different aspects of method performance, reducing reliance on a single measure and supporting a more balanced evaluation.

2.7. Ethical Considerations

In this research, humans did not participate as research objects and personal data were not used. The dataset did not contain student data, grades or personal identification information. All data contained only course structural information. Ethical approval was not required.

2.8. Generative AI Usage Disclosure

In this study, the generative artificial intelligence tool ChatGPT was used. It was used as one method for generating resource–course topic scores, as described in Section 2.4.3. In addition, ChatGPT was used for text editing, including grammar, spelling, punctuation and lexical error correction, as well as translation into the English language. In both cases, version 5.2 of the ChatGPT model was used. Generative AI was not used for any other purposes.

3. Results

In this part of the article, the results of comparing automated text similarity methods with expert-defined resource–topic mappings are presented. Several evaluation metrics were calculated, including mean absolute error (MAE), mean squared error (MSE), average difference, standard deviation of differences, and accuracy. In the analysis of the results, the focus is placed on MAE, MSE, and accuracy, as these metrics are the most informative for method comparison. Mean difference and standard deviation were used as supplementary metrics. However, they did not affect the relative ranking of the methods and are therefore not discussed in detail. Detailed results are presented in the following subsections.

3.1. Overall Comparison with Expert Evaluation

This section presents the aggregated results of all analyzed courses, comparing automated methods with expert evaluations. Metrics such as MAE, MSE, and accuracy were used to provide an overall performance assessment. The reported values represent averages across all analyzed courses.

Clear performance differences between the applied methods are observed. The LLM-based method achieved the best overall results, showing the lowest MAE and the highest average accuracy across courses, indicating the closest approximation to expert judgement. This performance is likely related to the model's ability to interpret contextual relationships between learning resource data and course topics rather than relying solely on word overlap.

In contrast, SBERT-based methods exhibit high error values and low accuracy. Moodle resource titles and descriptions are often short and generic, providing limited semantic information for embedding models. When texts lack context, semantic similarity becomes less reliable, resulting in weaker alignment with expert evaluations. Additionally, some course topics might contain Information and Communication Technology (ICT) area-specific terminology, which complicates the accurate embedding of those terms. If embedding models for each study field or even program existed, this could potentially improve the comparison results by providing better context representation.

Distance-based approaches using Euclidean distance demonstrate extremely poor performance across methods, suggesting a mismatch between sparse high-dimensional text vectors and distance-based metrics. In such representations, Euclidean distance can amplify small differences between vectors, leading to very large error values and near-zero accuracy. Meanwhile Jaccard and cosine distance based lexical comparison results are surprisingly high, showing the 2nd and 3rd result among the experiments. These distance metrics focus on similarity rather than absolute magnitude. Cosine distance measures the angle between vectors, capturing similarity in patterns while ignoring size, which is especially effective for text and embeddings. Jaccard distance measures overlap between sets, making it ideal for binary or sparse data where shared absence should not matter. Considering the different lengths of the compared texts, Jaccard and cosine distances were more suitable to reflect the topic rather than achieving a full text match.

Overall, the results presented in Table 3 suggest that methods capable of capturing contextual meaning are more suitable in this case, where the relationship between resource data and course topics is not always explicitly defined.

Table 3. Performance of automated methods across all courses compared to expert evaluation.

Method	Mean MAE	Mean MSE	Mean Accuracy (≤3)
LLM-based (ChatGPT)	1.52	31.22	88.09%
Lexical (Jaccard)	2.23	31.69	77.97%
Lexical (TF-IDF cosine)	2.64	50.96	77.18%
Lexical (TF-IDF Euclidean)	41.78	1747.00	0.00%
SBERT Cosine	21.18	695.11	8.70%
SBERT Euclidean	44.66	2005.55	0.00%

Considering that the LLM-based solution achieved the lowest mean absolute error and that the standard deviation across the analyzed courses was relatively small (0.56), the results indicate comparatively small deviations from expert evaluations within the analyzed dataset. The observed variation (approximately 0.4–2.6%) is smaller than the 5% step size used in expert scoring.

3.2. Performance of Lexical Similarity Methods

This section presents the results of lexical similarity methods. Three lexical techniques were analyzed: Jaccard similarity, TF-IDF cosine similarity, and TF-IDF Euclidean similarity. The performance of these methods is evaluated across individual courses.

Among the lexical approaches, Jaccard similarity shows the best performance across all courses, achieving lower error values and higher accuracy compared to TF-IDF cosine similarity. The strong performance of the Jaccard method suggests that direct term overlap remains useful in educational datasets, where course topics and resource titles often share key terminology. Its simplicity also makes it more robust when texts are short.

TF-IDF cosine similarity demonstrates slightly weaker performance but remains relatively stable across courses. This stability suggests that weighted lexical methods can serve

as reliable baseline approaches, particularly when similar terminology is used throughout a course.

In contrast, the TF-IDF Euclidean method performs poorly in all courses, exhibiting high error values and zero accuracy regardless of the number of learning resources or course topics. This indicates limited agreement with expert-defined mappings. The results also suggest that distance-based metrics are not well suited for sparse textual data and should therefore be applied cautiously in such tasks.

Overall, lexical methods demonstrate moderate but consistent performance, indicating that simple term-based approaches remain strong baselines for resource–topic alignment tasks. Detailed lexical similarity results for each course are presented in Table 4.

Table 4. Performance of lexical similarity methods.

Course ID	Method	MAE	MSE	Accuracy (≤3)
807	Jaccard	1.50	18.88	84.21%
807	TF-IDF cosine	1.72	29.98	84.12%
807	TF-IDF Euclidean	41.63	1733.48	0.00%
2858	Jaccard	2.09	29.10	82.17%
2858	TF-IDF cosine	2.35	44.25	79.72%
2858	TF-IDF Euclidean	41.73	1741.82	0.00%
3941	Jaccard	3.32	54.30	65.52%
3941	TF-IDF cosine	4.12	96.57	66.48%
3941	TF-IDF Euclidean	42.04	1771.92	0.00%
6821	Jaccard	2.02	24.47	80.00%
6821	TF-IDF cosine	2.36	33.03	78.41%
6821	TF-IDF Euclidean	41.72	1740.80	0.00%

3.3. Performance of Semantic Similarity Methods

This section presents the results of semantic similarity methods based on SBERT embeddings. Two approaches from this category were analyzed: SBERT with cosine similarity and SBERT with Euclidean similarity. The performance of these methods is evaluated across individual courses.

Semantic similarity methods perform worse than lexical and LLM-based approaches across all courses. SBERT cosine similarity exhibits high error values and low accuracy, indicating limited agreement with expert-defined mappings, while SBERT Euclidean similarity shows the poorest performance among all evaluated methods, with very high error values and zero accuracy.

One explanation for these results is the nature of Moodle resource data. Resource titles and descriptions are often very short and lack contextual information, providing limited semantic signals for transformer-based models and reducing the effectiveness of SBERT embeddings.

Another contributing factor is the structure of course topic titles, which frequently contain abstract or administrative phrasing rather than descriptive language. This weakens the semantic match between resources and course topics and may further explain the low performance of embedding-based models in this setting.

The findings suggest that the effectiveness of semantic similarity methods is strongly influenced by dataset characteristics, particularly text length and contextual clarity. Detailed semantic similarity results for each course are presented in Table 5.

Table 5. Performance of semantic similarity methods by course.

Course ID	Method	MAE	MSE	Accuracy (≤3)
807	SBERT cosine	15.93	411.24	15.61%
807	SBERT Euclidean	43.67	1912.40	0.00%
2858	SBERT cosine	17.28	450.10	10.49%
2858	SBERT Euclidean	43.89	1931.08	0.00%
3941	SBERT cosine	26.29	974.58	3.94%
3941	SBERT Euclidean	45.62	2095.56	0.00%
6821	SBERT cosine	25.21	944.53	4.76%
6821	SBERT Euclidean	45.47	2083.17	0.00%

3.4. Performance of the LLM-Based Method

This section presents the results of the LLM-based method (ChatGPT) in comparison with expert evaluations across individual courses. The LLM-based method demonstrates the strongest agreement with expert evaluations among all automated approaches, achieving the lowest error values and the highest accuracy when compared to lexical and semantic similarity methods.

The strong performance of this approach may be explained by its ability to interpret meaning and identify relationships between course topics and resource titles, even when direct textual similarity is limited. Unlike traditional similarity methods that rely mainly on word overlap or vector distance, LLMs can capture contextual information more effectively.

The findings suggest that LLM-based solutions can serve as useful tools for resource–topic alignment. Detailed LLM-based results for each course are presented in Table 6.

Table 6. Performance of the LLM-based method by course.

Course ID	MAE	MSE	Accuracy (≤3)
807	0.97	22.20	93.42%
2858	1.31	21.60	88.81%
3941	2.29	66.12	86.65%
6821	1.50	14.97	83.49%

3.5. Qualitative Analysis of Methods Agreement with Expert Mapping

A qualitative analysis was conducted using selected examples of resource–topic mappings to provide deeper insight into the observed performance patterns. These cases illustrate both agreement, disagreement, and non-alignment between automated methods and the expert-defined mappings.

The examples were intentionally selected as representative cases reflecting broader trends across the dataset. This approach enables clearer interpretation of alignment errors and methodological differences. For each automated method, the reported topic represents the highest-scoring match within the similarity matrix. The selected examples are presented in Table 7.

The results presented in Table 7 indicate that the effectiveness of similarity methods is closely related to how clearly learning resources reflect their associated topics. When the topic is explicitly stated in the resource title, all automated methods demonstrate strong agreement with the expert-defined mapping, as illustrated in the first and third examples.

In contrast, when key terminology is not directly included in the title, lexical similarity produces weaker alignment. This is evident in the second example, where the resource title does not explicitly mention MapReduce. As a result, the lexical method assigned the resource to a less relevant topic, whereas the SBERT-based approach successfully identified the correct mapping by utilizing contextual information from the resource description.

Table 7. Agreement, disagreement, and non-alignment.

Resource	Description	Expert	Lexical	SBERT	LLM
Distributed file systems	-	Distributed file systems and cache. NFS, AFS.	Distributed file systems and cache. NFS, AFS.	Distributed file systems and cache. NFS, AFS.	Distributed file systems and cache. NFS, AFS.
Midterm exam no 2	During the lecture implement one task from block A and one task from block B. The midterm is on MapReduce principles.	Map-reduce application for data processing	Distributed file systems and cache. NFS, AFS.	Map-reduce application for data processing	Distributed file systems and cache. NFS, AFS.
Spark and other frameworks	-	Spark usage for data analysis.	Spark usage for data analysis.	Spark usage for data analysis.	Spark usage for data analysis.
Homework	Task	-	-	Spark usage for data analysis.	-

Another important observation concerns resources that were not assigned to any topic by the expert. In such cases, most methods correctly avoided mapping the resource. However, the SBERT model generated a false-positive alignment, suggesting that semantic models may occasionally infer relationships that are not actually present.

Overall, these examples support the trends identified in the previous sections and provide additional insight into how alignment errors occur across different similarity methods.

4. Discussion

The results of this research show clear differences in how automated methods perform when mapping Moodle learning resources to course topics in comparison with expert-defined mappings. These differences make it possible to identify more suitable and effective methods for automated resource–topic alignment.

Taken together, these findings suggest that the effectiveness of text similarity methods depends strongly on dataset characteristics rather than on methodological complexity alone. Lexical approaches tend to perform better when terminology overlaps and texts are short, whereas embedding-based models require richer contextual information to operate reliably. In contrast, LLM-based methods appear particularly well suited for such situations due to their ability to interpret meaning beyond explicit textual similarity.

Among all evaluated approaches, the LLM-based method demonstrates the strongest agreement with expert evaluations across all courses. It achieves the lowest error values and the highest accuracy, indicating that large language models are well suited for scenarios where the available textual information is limited. This behavior is the most similar to expert judgment and may be attributed to the reasoning and contextual understanding capabilities of LLMs.

Lexical similarity methods show moderate performance. These approaches can partially reproduce expert-defined mappings, but their reliance on surface-level text overlaps limits their ability to capture deeper semantic relationships. As a result, their performance is lower than that of the LLM-based method, although they consistently outperform semantic embedding approaches.

Semantic similarity methods based on SBERT embeddings perform worst in this study. When applied to short and generic Moodle resource titles and descriptions, these methods produce high error values and low accuracy. This indicates that semantic embedding models are less effective in cases where textual information is limited and lacks sufficient contextual detail.

An important characteristic of the mapping results is the distribution of alignment values across the matrices. Both expert-defined and automated mappings, particularly

those produced by lexical and LLM-based methods, contain a high proportion of zero-valued cells. This reflects learning resources that are not aligned with specific course topics and partially explains the higher agreement observed for these methods compared to SBERT-based approaches.

From a practical perspective, these findings suggest that institutions seeking to implement automated evaluations similar to that examined in this study should carefully consider dataset properties when selecting similarity methods.

Several limitations of this study should be acknowledged. Moodle learning resource metadata is often short, general, and limited in semantic content, which negatively affects the performance of semantic similarity methods. In addition, the mapping matrices were created by two experts working independently. Although both experts had a strong understanding of course structure and quality assurance requirements, no formal inter-rater agreement metric was computed. As a result, the expert matrix should be interpreted as a structured expert benchmark rather than a statistically validated ground truth. This introduces a degree of subjectivity that may influence evaluation outcomes and the comparative performance assessment of automated methods. Furthermore, the LLM-based evaluation was conducted using the ChatGPT web interface, which does not allow full control over generation parameters. Although a fixed prompt was consistently applied to reduce variability, complete reproducibility cannot be guaranteed, and minor output variation across runs remains possible.

Despite these limitations, the findings indicate that automated methods can approximate expert evaluations when applied to realistic Moodle data. The results highlight the importance of selecting appropriate methods for effective resource–topic alignment and demonstrate the strong potential of LLM-based approaches for analyzing course structure and supporting quality assurance in digital learning environments.

Future research could address the mentioned limitations in several ways. First, incorporating additional data sources, such as the full content of learning materials rather than only resource titles and descriptions, could provide richer semantic context and improve the performance of automated alignment methods. Second, involving additional experts and formally assessing inter-rater agreement would strengthen methodological robustness and reduce potential bias in the evaluation framework. Third, future studies could employ Application Programming Interface (API)-based LLM implementations with controlled generation parameters to enhance reproducibility and enable systematic evaluation of output stability.

Author Contributions: Conceptualization, B.G. and S.R.; methodology, S.R.; software, B.G.; validation, S.R. and B.G.; formal analysis, B.G.; investigation, B.G. and S.R.; resources, S.R.; data curation, B.G.; writing—original draft preparation, B.G.; writing—review and editing, S.R.; visualization, B.G.; supervision, S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to institutional and privacy restrictions, as the dataset is based on internal university Moodle course resources and official study program documentation.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (OpenAI, version 5.2) for generating resource–course topic scores as described in Section 2.4.3, as well as for text editing, including grammar, spelling, punctuation, lexical corrections, and translation into English. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
API	Application Programming Interface
HTML	Hypertext Markup Language
ICT	Information and Communication Technology
LLM	Large Language Model
LMS	Learning Management System
MAE	Mean Absolute Error
Moodle	Modular Object-Oriented Dynamic Learning Environment
MSE	Mean Squared Error
NLP	Natural Language Processing
SBERT	Sentence Bidirectional Encoder Representations from Transformers
TF-IDF	Term Frequency-Inverse Document Frequency

References

1. Nataliia, L.; Liudmyla, I.; Mykhailivna, D.N.; Ivanna, V.; Olha, Y.; Ivanivna, B.S. Moodle innovation learning technology for medical education: From theory to practice. *J. Pharm. Res. Int.* **2021**, *33*, 245–260. [CrossRef]
2. Aljad, R.R. Analysis of development trends and experience of using LMS in modern education: An overview. *E-Learn. Innov. J.* **2023**, *1*, 86–104. [CrossRef]
3. Gamage, S.H.; Ayres, J.R.; Behrend, M.B. A systematic review on trends in using Moodle for teaching and learning. *Int. J. STEM Educ.* **2022**, *9*, 9. [CrossRef]
4. Al Abri, M.; Elhaj, A. Quality criteria for online courses development. *Int. Rev. Res. Open Distrib. Learn.* **2025**, *26*, 205–226. [CrossRef]
5. Lister, M. Trends in the design of e-learning and online learning. *J. Online Learn. Teach.* **2014**, *10*, 671–679.
6. Wongvorachan, T.; Bulut, O. The use of natural language processing in learning analytics. In *Advanced Learning Analytics Methods*; Springer: Cham, Switzerland, 2025; pp. 207–234. [CrossRef]
7. Lan, Y.; Li, X.; Du, H.; Lu, X.; Gao, M.; Qian, W.; Zhou, A. Survey of natural language processing for education: Taxonomy, systematic review, and future trends. *IEEE Trans. Knowl. Data Eng.* **2025**, *in press*. [CrossRef]
8. Allam, H.; Makubvure, L.; Gyamfi, B.; Kwadwo, G.; Akinwolere, K. Text classification: How machine learning is revolutionizing text categorization. *Preprints* **2024**. [CrossRef]
9. Madatov, K.; Sattarova, S.; Vičič, J. TF-IDF based classification of Uzbek educational texts. *Preprints* **2025**. [CrossRef]
10. Asudani, D.S.; Nagwani, N.K.; Singh, P. Impact of word embedding models on text analytics in deep learning environment: A review. *Artif. Intell. Rev.* **2023**, *56*, 10345–10425. [CrossRef]
11. Arafeh, S. Curriculum mapping in higher education: A case study and proposed content scope and sequence mapping tool. *J. Further High. Educ.* **2015**, *40*, 585–611. [CrossRef]
12. Margienė, A.; Ramanauskaitė, S.; Nugaras, J.; Stefanovič, P.; Čenys, A. Competency-Based E-Learning Systems: Automated Integration of User Competency Portfolio. *Sustainability* **2022**, *14*, 16544. [CrossRef]
13. Ako-Nai, F.; de la Cal Marin, E.; Tan, Q. A machine-learning based approach to validating learning materials. In Proceedings of the Computational Intelligence in Security for Information Systems Conference, Cham, Switzerland, 27 August 2023; pp. 306–315.
14. Ghosh, S.; Desarkar, M.S. Class specific TF-IDF boosting for short-text classification. In Proceedings of the Web Conference 2018, Lyon, France, 23–27 April 2018; pp. 1629–1637. [CrossRef]
15. Shyrokykh, K.; Girnyk, M.; Dellmuth, L. Short text classification with machine learning in the social sciences. *PLoS ONE* **2023**, *18*, e0290762. [CrossRef] [PubMed]
16. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Zhou, D. Chain-of-thought prompting elicits reasoning in large language models. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 24824–24837.
17. Bubeck, S.; Chandrasekaran, V.; Eldan, R.; Gehrke, J.; Horvitz, E.; Kamar, E.; Zhang, Y. Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv* **2023**, arXiv:2303.12712. [CrossRef]
18. Jurafsky, D.; Martin, J.H. *Speech and Language Processing*, 3rd ed.; Stanford University: Stanford, CA, USA, 2025.
19. Turney, P.D.; Pantel, P. From frequency to meaning: Vector space models of semantics. *J. Artif. Intell. Res.* **2010**, *37*, 141–188. [CrossRef]
20. Gomaa, W.H.; Fahmy, A.A. A survey of text similarity approaches. *Int. J. Comput. Appl.* **2013**, *68*, 13–18. [CrossRef]

21. Mu, W.; Lim, K.H. Modelling text similarity: A survey. In Proceedings of the ASONAM, Istanbul, Türkiye, 6–9 November 2023; pp. 698–705. [CrossRef]
22. Widianto, A.; Pebriyanto, E.; Fitriyanti, F.; Marna, M. Document similarity using TF-IDF representation and cosine similarity. *J. Dinda Data Sci. Inf. Technol.* **2024**, *4*, 149–153. [CrossRef]
23. Halim, J.; Lasut, D. Document plagiarism detection application using TF-IDF and cosine similarity. *Bit-Tech* **2024**, *7*, 202–213. [CrossRef]
24. Kowsari, K.; Jafari Meimandi, K.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text classification algorithms: A survey. *Information* **2019**, *10*, 150. [CrossRef]
25. Ahmad, F.; Faisal, M. A novel hybrid methodology for computing semantic similarity. *Int. J. Cogn. Comput. Eng.* **2022**, *3*, 58–77. [CrossRef]
26. Amur, Z.H.; Kwang Hooi, Y.; Bhanbhro, H.; Dahri, K.; Soomro, G.M. Short-text semantic similarity: Techniques, challenges and future perspectives. *Appl. Sci.* **2023**, *13*, 3911. [CrossRef]
27. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In Proceedings of the EMNLP-IJCNLP, Hong Kong, China, 27 August 2019; pp. 3980–3990. [CrossRef]
28. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Amodei, D. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
29. Kojima, T.; Gu, S.S.; Reid, M.; Matsuo, Y.; Iwasawa, Y. Large language models are zero-shot reasoners. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 22199–22213.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.