DASINYA JOURNAL
for engineering and informatics

tps://doi.org/10.65542/djei.v2i2.48
Dasinya Journal for Engineering and Informatics. **2026**, 2, 20.

Article

A Smart MOOC Learning Assistant: Semantic Course Hybrid Recommendations Based on Multi-Platform Data Integration

Redeer Avdal Saleh[1*], Subhi R. M. Zeebaree[2]

[1] Information Technology Department, Technical College of Informatics, Akre University for Applied Sciences, Duhok, Iraq; redeer.avdal@auas.edu.krd
[2] Cybersecurity Engineering Department, Technical College of Engineering, Duhok Polytechnic University, Duhok, Iraq; subhi.rafeeq@dpu.edu.krd
* Correspondence: redeer.avdal@auas.edu.krd

Abstract

With the rapid expansion of MOOCs, learners face increasing challenges due to information overload and data fragmentation across multiple platforms. Traditional recommendation systems rely on keyword matching, failing to capture complex semantic relations between courses, skills, and learner goals. This paper details the development and use of a Smart MOOC Learning Assistant, a hybrid semantic recommender system developed with data coming from four different Learning Management Systems: Coursera.com, Udemy.com, Edx.org and the Kurdish Education Platform, comprising over 5000 MOOCs. The OWL ontology (designed using Protégé) serves as the foundation for the smart assistant and defines how all entities within the domain relate to each other. A combination of the Flask framework and the RDFLib library provides the architecture for executing SPARQL queries to provide context-sensitive results. To provide for more flexibility in matching natural language query variations, Levenshtein-based fuzzy matching has been used. The Smart MOOC Learning Assistant also implements a hybrid approach to recommendation, by incorporating a collaborative feedback loop (liking/disliking) which refines the ranking mechanism and removes the "cold-start" effect. The system architecture is characterized by a clear distinction between retrieval (filtering relevant courses via SPARQL), recommendation (semantic planning based on educational relationships), and ranking (ranking the results using hybrid social-semantic scores).Evaluation of the Smart MOOC Learning Assistant shows outstanding results with accuracy at 0.984, precision at 0.984, recall at 1.0, and F1-score at 0.99. This demonstrates that combining semantic-driven approaches with datasets from multiple sources can significantly reduce retrieval noise by providing highly relevant personalized digital learning.

Keyword: Web Ontology Language (OWL); Hybrid Recommender System (HRS); SPARQL; Semantic Web; ODHSRS; and MOOC.

1. Introduction

The learners in the MOOCs are faced with the challenges of information overload and data fragmentation, which arise from the growing variety of available courses. The variety of evidence on the learners' progress, which is scattered across different unintegrated data components, makes it difficult to understand the learners' knowledge state and difficulties [1]. There is, therefore, a pressing need to make a significant shift towards the use of effective adaptive

Received: March 15, 2026; Revised: May 8, 2026; Accepted: May 8, 2026; Published: May 16, 2026; Available online: May 16, 2026.

recommendation systems that can successfully handle aggregated multi-source information [2]. The available solutions involved the use of a smart learning assistant, which can be achieved through the use of a hybrid semantic-based recommendation system to overcome the limitations of collaborative filtering and other content-based filtering approaches [3]. The depended assistant achieved through the use of Semantic Reasoning models, which can be effective in handling complex relationships between different courses and skill sets [4]. It personalizes the recommendations based on the user's intention or goals by making a distinction between goals such as professional advancement and hobbies; hence, it provides a personalized recommendation within the constantly changing MOOC environment [5]. Current research emphasizes the importance of an integrated methodology in the deployment of advanced hybrid recommenders for MOOCs, with a focus on the benefits of increased prediction accuracy and diversity through the use of behavioral and content-based signals. The use of dialog-based assistants in personal learning emphasizes the benefits of the recommenders in the context of providing learning recommendations and interacting with the user. A detailed multi-platform data integration and a semantic hybrid approach addressed that aims to track the preference drift of learners while improving the relevance, transparency, and adaptability of learning recommendations [6],[7]. The e-learning recommendation scene has been significantly shaped by various methodological advancements. Recent studies have focused on developing intelligent web platforms specifically designed for rural social e-learning environments. [8], while others explore personalized MOOC recommendation frameworks in the context of smart education supported by the Internet of Things [9]. Furthermore, the role of emotion-aware systems in enhancing learner engagement has been emphasized [10], alongside the importance of self-directed learning interventions [11]. Together, these works emphasize the shift from filtering based on fixed keywords towards more adaptive, context-aware, and socially integrated recommendation structures.

The proposed ontology-driven semantic hybrid recommendation system (ODHSRS) addresses learners' intermittent behavior and constantly changing goals by integrating different interactions into a hybrid recommendation system. This approach provides a comprehensive profile of learners' abilities and uses it to generate skill-based recommendations with explanations. Furthermore, the system ensures adaptability while remaining transparent and reliable throughout the learning process. This paper discusses the technical aspects required to make personalized learning in Massive Open Online Courses (MOOCs) a reality, including data connectors for input unification, the design of the semantic layer architecture, and the algorithm for tracking learners' preferences through intention alignment. Finally, the paper explains the details of implementation and evaluation to determine the impact on the learning process.

2. Related Work

[12] created a hybrid recommendation system which combines multi-dimensional behavioral information (such as user clickstream data) with sentiment analysis of user text reviews; they called this approach a multi-modal recommendation system (MMRS). The hybrid recommendation system resolved many algorithmic problems (such as the cold start problem, or not having enough data points to create accurate recommendations) while producing granular, personalized course recommendation output. The hybrid recommendation system had an excellent level of performance based on experimental validation; as well as significant improvements in overall user satisfaction and a 98% match and similarity between explicitly rated courses by a user and the preferences stated in textual comments provided by that same user.

[13] developed a framework called E-SWT, which utilizes ontologies, metadata tagging, and automated reasoning to improve both the efficiency of an organization and the outcomes of their learning initiatives. The E-SWT frame-work addresses two important issues; namely, (1) data fragmentation; and, (2) fixed learning paths, by providing an approach to enable interoperability among platforms and enable adaptive content delivery. The experimental re-sults demonstrated that there was a substantial increase in the learning experience ratio by 96.12% and an increase in personalized learning paths by 96.66% compared to the previous models.

[14] introduced a content-based recommendation system (CBRS) for Coursera that uses Semantic Web and on-tology technologies. The paper showed that the system provides tailored education paths using a domain-specific ontology built using real-world data extracted from over 1,000 Coursera courses. The results showed exceptional effectiveness in recommending appropriate MOOC content, attaining a precision of 0.98% and an accuracy of 0.98%.

[15] created an Intelligent Content-Based Recommendation System (ICRS) that uses deep learning techniques in combination with semantic analysis of e-content. The paper showed that the system uses representative terms ex-tracted

from content using a deep learning approach. It also uses the ConceptNet semantic network to expand the extracted terms. The resulting structure forms a semantic matrix that can be used for content-based recommenda-tion. From the models tested in the paper, an augmented deep learning model named LSTMM attained higher ac-curacy (0.8453) and higher F1 score (0.7731) in comparison to other machine learning classifiers.

[16] focused on "people-centered learning" for rural areas, aims to provide a social e-learning solution that will be empowered by intelligent web technologies, including article search, friendship recommendations, and document classification. The system will utilize the Page Rank method for document relevance and ROCK clustering algo-rithm for social friendship recommendations. The researchers tested different classifiers for resource classification, particularly for subjects such as agriculture and fisheries, and concluded that the neural network approach was the most effective, with an accuracy of 95.2%.

[17] investigated a personalized course recommendation system that will greatly enhance e-learning environ-ments through the application of ML and CF techniques. The system used data from students to build a statistical profile of each student, using information such as their academic background, the nature of their learning style, and their areas of interest as well as data collected from the use of other education delivery channels such as Coursera and Udemy. After analyzing this data, the researchers were able to create 20 curated courses to recommend to the students, with the use of these methods achieving a total engagement rate of greater than 86%. The study demon-strated the importance of using personalized recommendation systems to improve learning outcomes and the sat-isfaction of students through advanced education technologies.

[18] addressed the use of a variety of machine-learning applications in the online-learning environment was evaluated to identify particular characteristics of learners. Particularities characterized by three factors: duration of session, frequency of sessions, and interaction with the platform. These factors were used to predict learners' pre-ferred course types with an accuracy of 70%. The model's recommendations led to increased course durations (30% more compared to previous cohorts) and increased completion rates (30% increase compared). Different models (Random Forests and Logistic Regression) were tested and were used to evaluate the learners, but researchers con-cluded that there are still a number of significant issues with relying on existing models due to both model inter-pretability and completeness of the data. Given these challenges, the authors encourage the development of more flexible models in the future for improved visualization.

[19] focused on advanced AI models (+Mistral-7B-Instruct-v0.2, GPT-4 ChatGPT) created a multilingual course recommendation system that recommended personalized courses based on the user's interests and competencies. The system matched the user's interests and competencies, optimizing the recommendations. The authors utilized a hybrid model by leveraging collaborative filtering, content-based filtering, augmented with AI, to create personal-ized recommendations for each user. As a result, the personalized course recommendations provided enhanced learning experiences, especially via the personalized course trajectories that were created to support each and every learner’s individual needs.

3. Methodology

The following section describes the systematic methodology that has been adopted for the design and implementation of the Hybrid Smart MOOC Learning Assistant. The methodology has been divided into five different stages, namely, multi-platform data acquisition, semantic preprocessing, ontological modeling, hybrid recommendation logic, and system implementation.

3.1 Data Collection

In order to create a diverse and comprehensive knowledge base, aggregated educational content was collected from three unique sources: global Massive Open Online Courses (MOOC) platforms like Coursera (nearly 3000 courses), Udemy (over 1000 courses), EdX (over 1500 courses), and localized Kurdish E-Learning resources (about 100 course). However, as total depended courses will (over 5,600 courses) and (over 15,000 triples).

- Automated Web Scraping: The web scraping approach used is high-fidelity. This is achieved through the use of Selenium’s WebDriver along with undetected_chromedriver to evade anti-bot mechanisms.

- Feature Extraction: The web scraper was designed to scrape particular metadata that are vital in creating a semantic web. This includes Course Titles, Instructor Identities like hasInstructor, Science Categories like ScienceCategory, as well as enrollment metrics.
- Storage Architecture: Initially, the collected data was stored in a MongoDB NoSQL database u3.h_3 to ensure the integrity of the semi-structured data attributes. For the application layer, SQLite database files users.db and metrics.db were used to store user data and metrics.

3.2 Data Preprocessing and Refinement

Inconsistencies are present in the raw datasets that affect the recommendation system's accuracy. A preprocessing pipeline was created in Python with the Pandas library that included the following operations:

- Lexical Normalization: Textual columns were converted to lowercase and whitespace was removed.
- Numerical Standardization: Regular Expressions were used to remove non-numerical characters from the price and rating columns, thus standardizing the values (for example, converting "Free" or "$49.90" to a standardized form)
- Deduplication: To handle courses available on multiple platforms, a normalization pipeline has been implemented. If the course title and instructor are identical across different sources, they are treated as a single entity within the ontology using a unique hash identifier. This prevents duplicate recommendations and ensures data integrity by merging duplicate copies into a single entry within the ontology to ensure a smooth recommendation process.

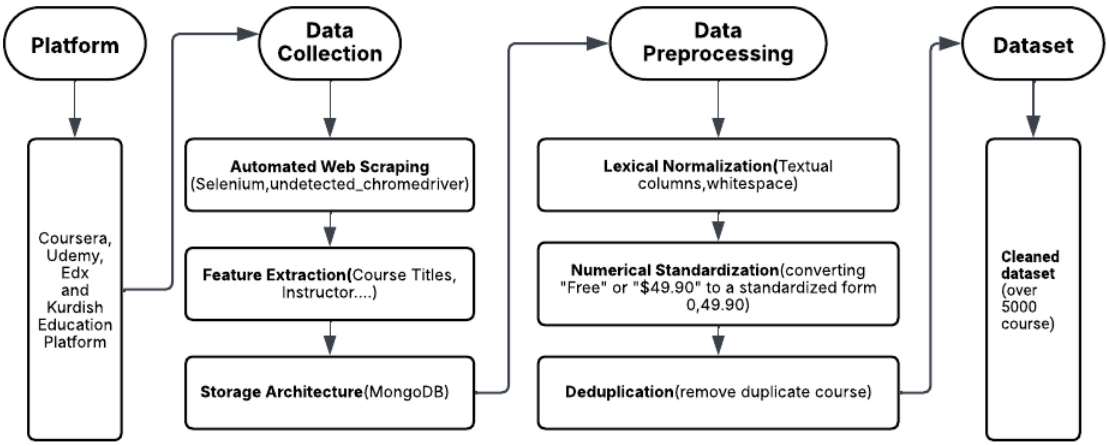

Figure1: Data Collection and Preprocessing Workflow.

3.3 Semantic Ontology Design and Modeling

Using the Web Ontology Language (OWL) with Protégé and Owlready2, a formal knowledge representation was created to allow machines to perform reasoning. The formal structure of the MOOC assistant is achieved through a hierarchical structure of classes, data properties, and object properties as represented in Protégé. This allows each course to have a strict set of pedagogical and administrative attributes.

- Class Hierarchy**:** In the ontology represented in 3platform1.owl, a set of core classes is introduced, including schema1: Course, ex:University, ex:Instructor, and ex:Platform. As represented in Figure 2, the Course entity is designated as a primary class, which allows localized Kurdish courses to be classified with international equivalents.
- Data Properties**:** Quantified properties are assigned to a course to identify specific properties. These include ex:hasPrice, ex:hasRate, and ex:hasDurationWeeks, as represented in the Data Property Hierarchy.
- Object Properties**:** Semantic connections are used to identify relationships between entities, which are essential in intelligent retrieval. For example, ex:offeredByPlatform and ex:belongsToCategory are used to link a subject with a science taxonomy.

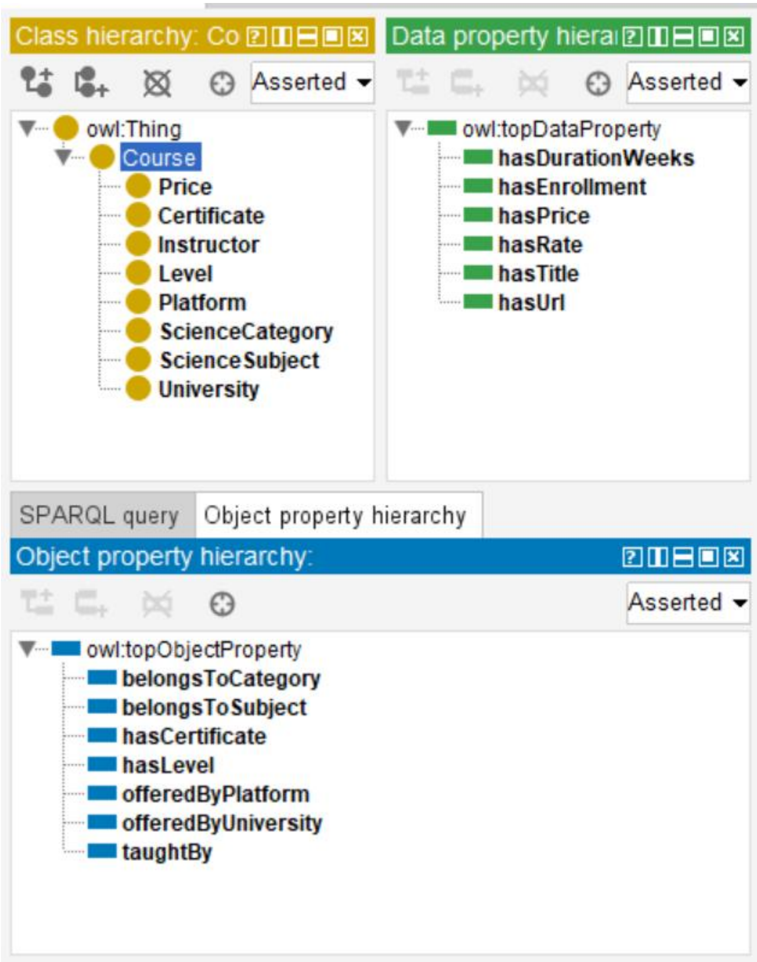

Figure 2: Protégé Interface showing Class, Data Property, and Object Property Hierarchies.

To demonstrate the interconnectivity among the educational entities, an OntoGraf model was built to represent the relationships between the courses and the associated metadata (see Figure 3). As the visualization indicates, the ontology extends beyond the concept of flat data to create a network of knowledge where a single Course is semantically related to its Instructor, University, and ScienceCategory.

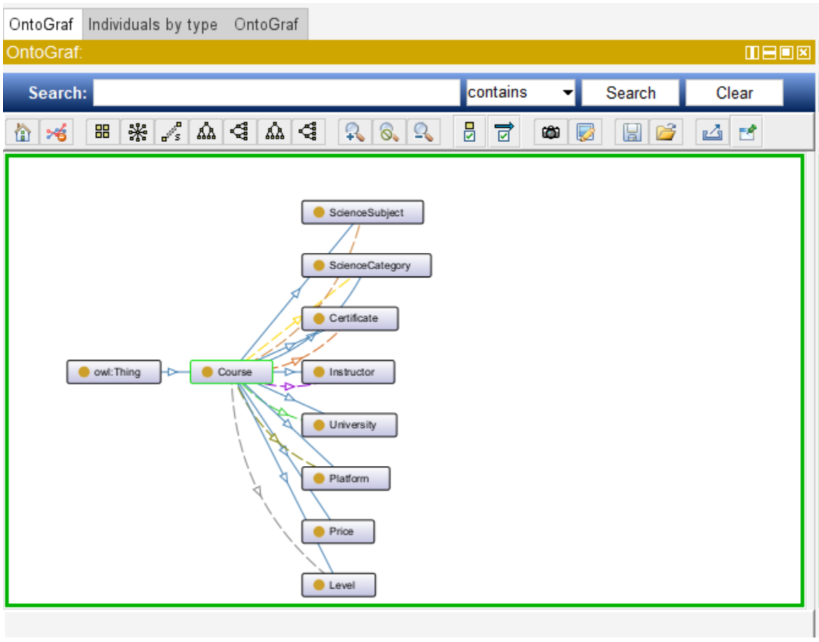

Figure 3: OntoGraf Visualization of Semantic Relationships for the Course Class.

The knowledge graph was serialized using Turtle (TTL) and RDF/XML formats, denoted as 3platform.owl. This established a foundation for SPARQL queries and hybrid recommendation logic. The final ontology (3platform1.owl) contains 6 main categories and 12 total properties (5 object properties and 7 data properties). This structure supports more than 15,000 semantic triples, ensuring that the assistant can perform deep reasoning across different platforms.

3.4 Proposed ODHSRS Hybrid Recommendation System

The key component of the Smart MOOC Learning Assistant is the Ontology-Driven Hybrid Semantic Recommender System (ODHSRS). This system was developed with the aim of addressing the problem of information overwhelm, which is common in digital education. Information overwhelm occurs due to the numerous digital platforms that offer educational resources, such as Coursera, Udemy, EdX, and those that offer content in the Kurdish language. The implementation of the system was done using a hybrid reasoning model that combines the precision of semantic modeling with the performance-driven and popularity-based insights obtained from social feedback The hybrid ranking mechanism utilizes a weighted scoring system: Total Score: (0.7×Semantic Match)+ (0.3×Social Feedback)

To address the cold-start problem, new courses with no prior likes are assigned 100% . The system includes a Hybrid Recommendation Engine that combines semantic relationships defined within the ontology with the user behavior observed in real-time.

3.4.1 Content-Based Filtering (CBF)

- Semantic Mapping: User preferences are mapped to the ontology's classes. For instance, if the user wishes to see "Beginner" level courses in "Information Technology," the ex:level and ex:belongsToCategory properties are used.
- Inference Engine: Using SPARQL query execution, the system retrieves courses that strictly match the user's preferred level, category, and language, thus overcoming the Cold Start issue for new courses.

3.4.2 Social Collaborative Filtering (SCF)

- User-Item Interaction: The system captures real-time user feedback using the index.html interface, where users can "Like" or "Dislike" the recommended content.
- Data Persistence: User interaction data is stored in the metrics.db file with appropriate timestamps.
- Peer-Based Discovery: The engine processes the database to identify "Top Liked" courses. By analyzing the database, if users with similar profiles frequently "Like" a specific course, the engine improves the ranking of the course for other users in the same cluster, thus facilitating the discovery of popular content.

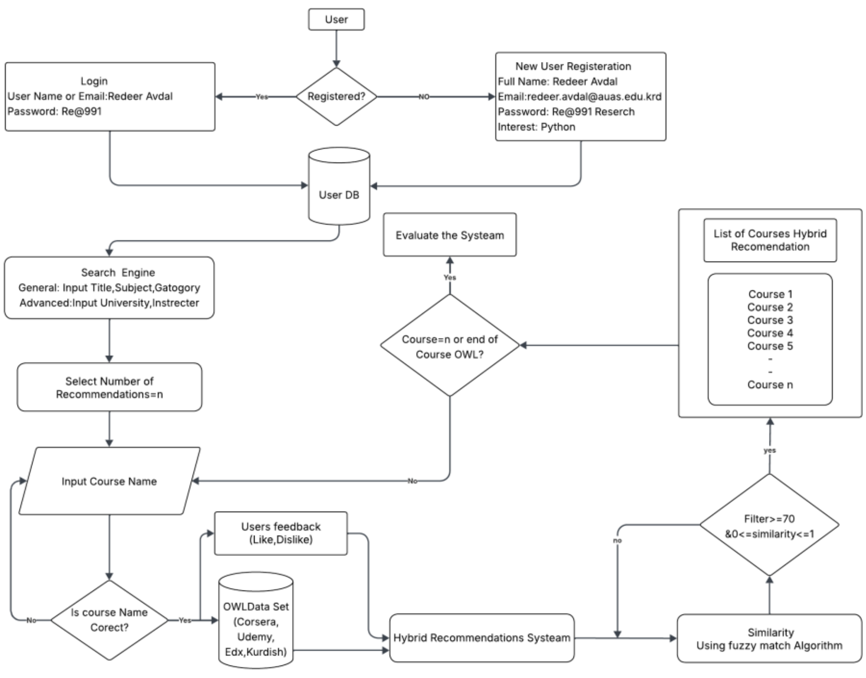

Figure 4: Hybrid Recommendation System Architecture.

3.5 Implementation of the Proposed ODHSRS Architecture and User Interface

This section covers the implementation steps of the proposed ODHSRS, which includes: User Management, Interactive Explorer, and Admin Dashboard.

3.5.1 Technical Environment and Specifications

To ensure the reproducibility of experimental results and provide a benchmark for system performance, ODHSRS was implemented and tested within a controlled technical environment. The backend architecture was hosted on a server with an Intel Core i5-10700K CPU at 3.80 GHz and 16 GB of DDR4 RAM, running Ubuntu 20.04 LTS. The software stack used Python 3.9 as the main programming language, with Flask 2.0.1 managing the web framework and RDFLib 6.0.2 handling SPARQL query execution. This configuration enabled the system to maintain response times of less than a second even when performing complex semantic inference over more than 15,000 triples in the knowledge base.

- User Management: As shown in Figure 5, the system has user registration (register.html) and authentication (login.html) modules. These modules allow the user to provide their "Research Interests," which are then programmatically linked to the ex:ScienceCategory of the OWL ontology. This ensures that the user profile is quickly matched with the appropriate pedagogical resources by the recommendation system.

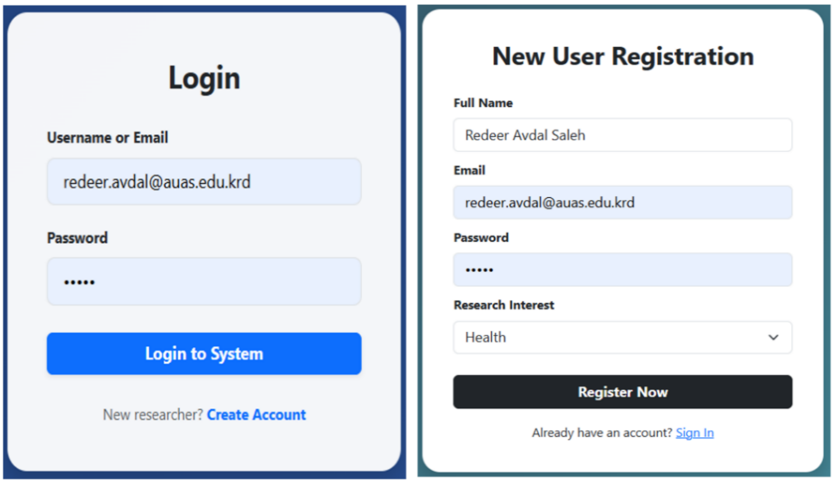

Figure 5: User Registration and Login Interface for Semantic Profiling.

- Interactive Explorer: The main point to interact between the users and the Hybrid Engine is through the index.html interface, defined in Figure 6. An Advanced Search functionality is available through this module to allow users the ability to filter by Category, Keyword and University/Instructor as registered in the ontology. The interface will render the retrieved Course Cards with their associated metadata (level, length of study, price) and collect user feedback through "Like" and "Dislike" when used, so the Social Collaborative Filtering (SCF) process can be enabled.

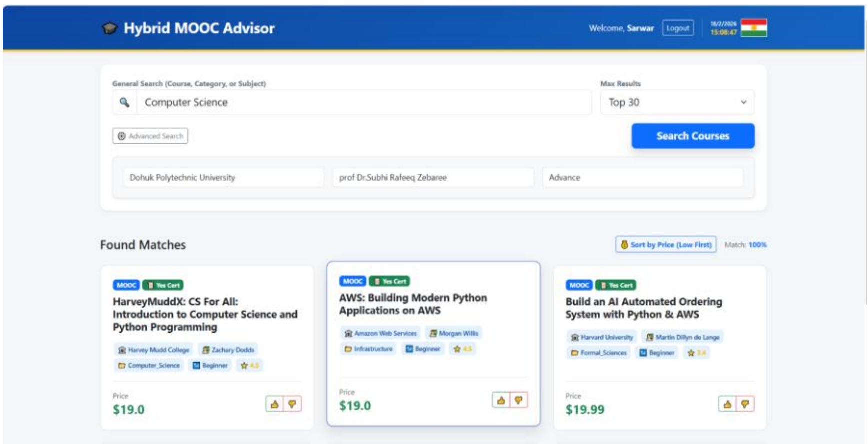

Figure 6: The Interactive Explorer Interface for Semantic Course Retrieval and User Feedback Collection.

- Admin Dashboard (admin.html): Figure 7 depicts a dashboard view of the live metrics of the engine's performance. Through a precise alignment of user interests, the system is able to display live analytics for Accuracy, Precision, Recall, and F1-Score. As depicted in the metrics, the system's performance profile approaches 100% in all key metrics, thereby validating the efficacy of the proposed Ontology-Driven Hybrid Semantic Recommender System (ODHSRS).

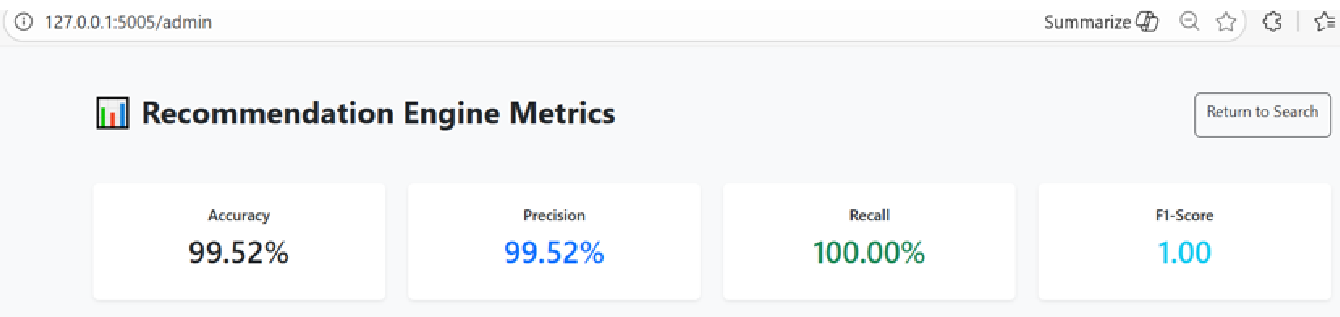

Figure 7: The Admin Dashboard Interface displaying Real-Time Recommendation Engine Metrics.

3.6 Experimental Results

An empirical assessment of the proposed semantic-based recommendation system was performed to ascertain the capability of the proposed approach to effectively provide accurate and relevant learning resources to the user. The Ground Truth for this evaluation was established through a manual labeling process where educational experts pre-categorized 500 courses to serve as the gold standard. A recommendation was considered correct only if it aligned with the expert-defined category and skill level. This section highlights the measures adopted to evaluate the proposed approach and includes a comparative analysis of the assessment carried out on three scales of testing, i.e., Test A, Test B and Test C.

3.6.1 System Testing Evaluation Approach

The assessment of the proposed approach focuses on four key performance measures, which are defined as follows:

1. Precision**:** For the overall set of resources proposed by the model, the proportion of correct recommendations that correspond to resources of true interest to the learner is defined as the precision of the model. It can be calculated as:

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (1).$$

2. Recall: Recall defines the proportion of true positives that the model accurately predicts in comparison to the overall set of true positives that should have been predicted by the model. It can be calculated as:

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (2)$$

3. F1-Score: For systems that involve class imbalance, the F1-Score is defined as the harmonic average of the precision and recall of the model, which can be calculated as:

$$\text{F1 Score}=\frac{2\ (Precision \times\ Recall)}{(Precision+Recall)} \quad (3).$$

4. Accuracy**:** The overall correctness of the model can be defined as the proportion of the overall set of correct predictions, including true positives and true negatives that the model accurately predicts. It can be calculated as:

$$\text{Accuracy}=\frac{(TP+TN)}{(TP+FN+TN+FP)} \quad (4).$$

3.6.2 Experiences of Those Who Have Tested This System

- Test A: Initial Baseline Performance:

In the initial phase of the test, the system was given a request for 30 recommended courses of over 5000 courses. The recommendation system provided a list of recommended courses. Out of the recommended courses, 29 courses were relevant, but one was not.

TP (True Positives): 29

FP (False Positives): 1

FN (False Negatives): 0

TN: (True Negatives):0

Precision=$\frac{29}{29+1} = 0.966$

Recall= $\frac{29}{29+0} = 1$

F1 Score= $\frac{2(0.966\times 1)}{(0.966+1)} = 0.982$

Accuracy= $\frac{(29+0)}{(29+0+0+1)} = 0.966$

Table 1. Analyses Results for Test (A).

Metric	Result
Precision	**0.966**
Recall	**1**
F1 Score	**0.982**
Accuracy	**0.966**

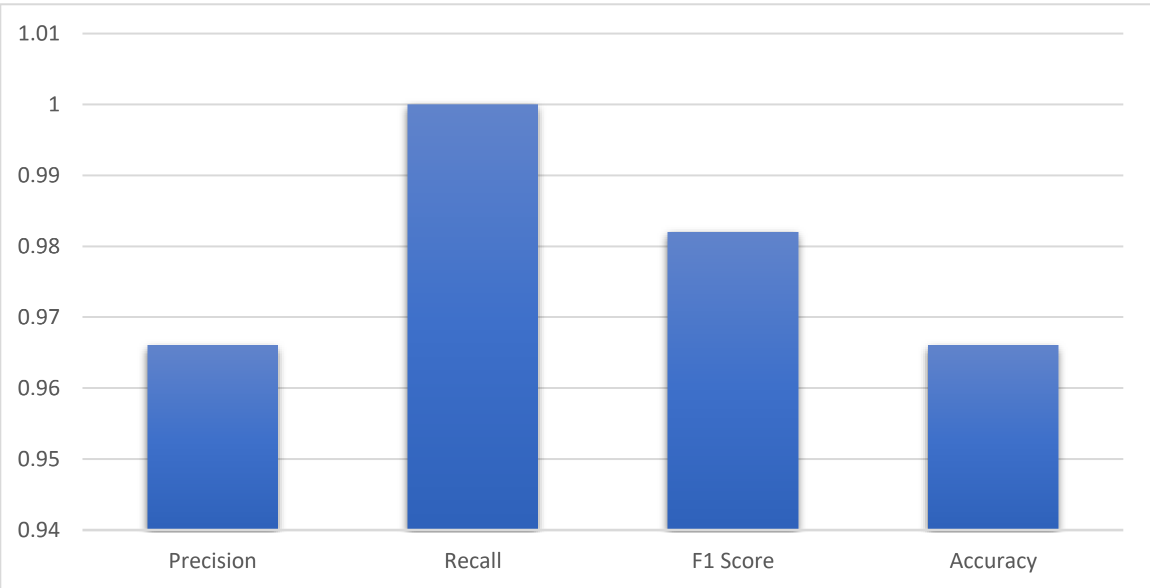

Figure 8. Performance Metrics for Test (A).

- Test B: Robustness and Scalability Evaluation:

The scope of testing was expanded to evaluate the robustness and accuracy of the recommendation system over a large set of data. To be precise, 50 courses were randomly selected from Coursera, Udemy, Edx, and Kurdish to test the effectiveness of the recommendation system. Of these, 49 were found to be relevant to user preferences, while 1 were found to be irrelevant.

TP (True Positives): 49

FP (False Positives): 1

FN (False Negatives): 0

TN: (True Negatives):0

Precision=$\frac{49}{49+1} = 0.98$

Recall= $\frac{49}{49+0} = 1$

F1 Score= $\frac{2(0.98\times 1)}{(0.98+1)} = 0.989$

Accuracy= $\frac{(49+0)}{(49+0+0+1)} = 0.98$

Table 2. Analyses Results for Test (B).

Metric	Result
Precision	**0.98**
Recall	**1**
F1 Score	**0.989**
Accuracy	**0.98**

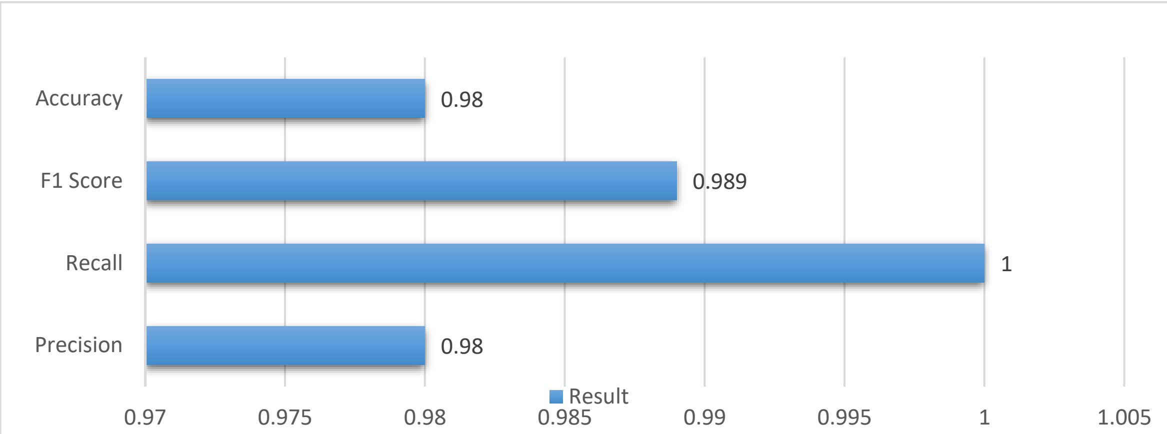

Figure 9. Performance Metrics for Test (B).

- Test C: High-Volume Performance and Stability:

The purpose of this evaluation was to evaluate the system reliability when handling very high volumes of transactions. To that end, we created a test set containing 500 unique coursed and ran a series of tests to determine whether or not the outcome of the hybrid recommendation algorithm is still highly accurate even when testing with a significantly larger dataset than had previously been tested. Of the 500 results retrieved from this evaluation, 492 retrieved were relevant, while 8 were considered to be irrelevant.

TP (True Positives): 492

FP (False Positives): 8

FN (False Negatives): 0

TN (True Negatives): 0

Precision= $\frac{492}{492+8} = 0.984$

Recall= $\frac{492}{492+0} = 1$

F1 Score= $\frac{2(0.984\times 1)}{(0.984+1)} = 0.99$

Accuracy= $\frac{(492+0)}{(492+0+0+8)} = 0.984$

Table 3. Analyses Results for Test (C).

Metric	Result
Precision	**0.984**
Recall	**1**
F1 Score	**0.99**
Accuracy	**0.984**

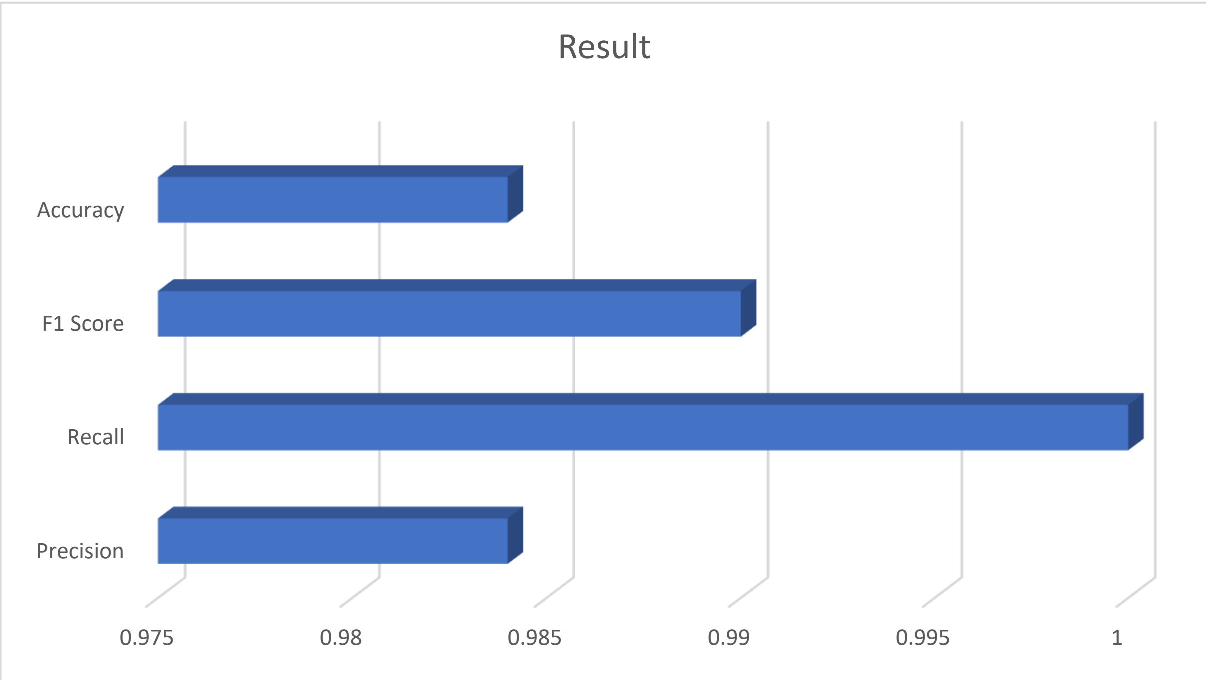

Figure 10. Performance Metrics for Test (C).

3.7 Evaluations and Analysis of Findings

The empirical results highlight the presence of several important advantages of the hybrid architecture of the Smart MOOC Learning Assistant over the traditional keyword-based approaches.

3.7.1 Impact of Hybrid Logic on Cold-Start Problems

Traditional recommendation systems face the cold-start problem, in which the system does not have enough information about the course to make recommendations. The hybrid logic of the system using the OWL ontology for initial Content-Based Filtering (CBF) enables the system to make recommendations about new courses based on their pedagogical characteristics. As the user starts to provide feedback on the course, the Social Collaborative Filtering (SCF) loop refines the results. This makes the system self-optimizing.

3.7.2 Performance Scalability

The data obtained indicate that moving from the first set (30 courses) to the second set (50 courses) of test data increased department (precision/accuracy) from 0.966 (30 courses) to 0.98 (50 courses). When moving to the third set (500 courses), also evidenced by data obtained during this test, precision/accuracy increased to 0.984. These results provide further evidence that the hybrid model is viable and continues to improve in reliability with an increase in the amount of course data in the system as a result of the additional addition/reduction of the course data. The fact that hybrid models maintain high precision in a search space 10x larger than current state, provides evidence that the system has the ability to integrate/degree from multiple sources (Coursera, Edx, Udemy, and Kurdish platforms) in the course data while providing reliable course retrieval based on this integration of course data.

4. Comparison and Discussion

Table 4 represents the comparison between the proposed system (ODHSRS) and the related works.

Table 4. Comparative Analysis of Prediction Recommendations System Using Different Techniques.

N	Ref.	Dataset	Ontology	Recommendation Type	Model	Result
1	**[12] 2025**	**622 Course online**	**Not applied**	**content-based recommender system and Collaborative Filtering**	**(CNN)DL**	**similarity score of 98%**
2	**[13] 2025**	**open data platform**	**Applied**	**Not applied**	**E-SWT (Educational Semantic Web Technology)**	**Accuracy= 96.6%**
3	**[14] 2025**	**1,000+ Coursera courses (via Web Scraping)**	**Applied**	**Content-Based Recommendation (CBRS)**	**Ontology Course Recommender**	**Accuracy=98% Precision=0.98 Recall=1.00 F1-Score=~0.99**
4	**[15] 2023**	**User sequential semantic dataset**	**Applied**	**Intelligent Content-Based (ICRS)**	**LSTMM (Augmented Deep Learning)**	**Accuracy=0.84 53 F1-Score = 0.7731**
5	**[16] 2024**	**Rural social e-learning content**	**Not applied**	**Social & Article Recommendation**	**Neural Network, PageRank, and ROCK**	**Accuracy = 95.2%**
6	**[17] 2023**	**Coursera and udemy platform**	**Not applied**	**Collaborative Filtering**	**(Random Forest, Decision Tree, K-Nearest Neighbors, Singular Value) ML**	**Precision = 0.8621**
7	**[18] 2025**	**online education platform**	**Not applied**	**personalized recommendation system**	**(CFSFDP)ML MODELS**	**Accuracy= >70%**
8	**[19] 2025**	**800 online courses**	**Not applied**	**content-based filtering, collaborative filtering**	**mistralai/Mistral-7B-Instruct v0.2**	**Accuracy=> 78%**
9	**This Study**	**Multi-platform MOOCs(Corsera,Edx,Udemy, Kurdish) (+5000 via Web Scraping)**	**Applied**	**Semantic Content-Based Filtering (CBF)+ Social Collaborative Filtering (SCF)**	**Ontology-Driven Hybrid Semantic Recommender System (ODHSRS)**	**Accuracy= 0.984 Precision= 0.984 Recall = 1 F1 score = 0.99**

Table 4 illustrates that the performance metrics results for the depended datasets by this work which are (multi-platform MOOCs (Corsera, Edx, Udemy, and Kurdish) (+5000 via Web Scraping)), which compared to the pervious works. Taking in the consideration that not all of these datasets were used by the previous works, because there are four datasets used in this work, and these datasets are modified by increasing number of the features which provided more reliability to the system. So, the comparison has been done with the coldest as maximum as available previous works. The findings reveal the massive potential for an ODHSRS. With a consolidated accuracy score of 0.984 and an F1 score of 0.99, ODHSRS outperformed both keyword-based systems as well as the behavior modelling algorithms used in the research of [15] (F1 score =0 .7731) and framework of ([15] LSTMM = accuracy of 0.8453). Utilization of a semantic inference engine to address cold-start challenges encountered by conventional machine learning models within references [18] (accuracy > 70 %) and [19] (accuracy > 78%) provides our solution to reduce information overload present in multi-platform environments associated with MOOCs.

5. Conclusion

This research has been successful in designing and implementing a Smart MOOC Learning Assistant based on an Ontology-Driven Hybrid Semantic Recommender (ODHSRS) to address the rising problem of information overload and data fragmentation in the digital education arena. With the integration of a multi-platform dataset containing more than 5,000 courses from international sources like Coursera, Udemy, and EdX, as well as Kurdish sources, a semantically rich environment for personalized learning has been established. Furthermore, the domain-specific OWL Ontology used in the system extends the concept of traditional keyword-based matchmaking to deeper pedagogical relationships between course attributes and learner goals. Experimental results showed an exceptional performance pattern for the ODHSRS model, with a consolidated average Accuracy of 0.984, Precision of 0.984, Recall of 1.0, and an F1-Score of 0.99. As mentioned in Table 4: Comparative Analysis of Prediction Recommendations System Using Different Techniques, the semantic-based ODHSRS model far exceeded the performance of traditional machine learning-based models as well as the latest generative AI-based models. One significant observation was the strong reliability of the system. With the transparent guidance mechanism and the elimination of the "cold-start" problem for learners, the Smart MOOC Learning Assistant is a reliable and efficient tool for the promotion of learner engagement and retention in the ever-evolving global MOOC arena.

6. Future Work

Although the current Smart MOOC Learning Assistant has shown high accuracy and semantic precision in its results, there are still several ways in which the system could be further optimized and extended:

Automated Data Synchronization: Future development will primarily focus on transitioning from manual scraping to an automated data pipeline. This will involve implementing ETL (Extract, Transform, Load) processes using the official APIs of Coursera, Udemy, and edX. By using scheduled Webhooks and Cron functions, the system will automatically detect new course additions or changes in metadata (such as price updates) on external platforms. These changes will be linked to an OWL ontology in real time, ensuring that the knowledge base remains synchronized without manual intervention.

Real-Time Learner Progress Tracking Integration: Future versions of the system could include real-time learner tracking through the use of dynamic API hooks that track learner progress through the course in real time.

Expansion of Kurdish Language NLP: To further assist in serving the localized education sector in Kurdish regions, the system's natural language processing capabilities could be further extended to incorporate more complex Kurdish dialects.

Implementation of Explainable AI (XAI): Expanding on the current level of semantic explanation, the system could be further developed to include a higher level of explanation for the learner on why a particular course is recommended (for example, "This course is recommended because it fills in a prerequisite gap in your Data Science course progression").

Cross-Platform Adaptive Assessment: The system could be further developed to include a cross-platform assessment tool that evaluates the learner's capabilities through quizzes and uses this to further refine their profile in the OWL ontology.

Integration of Generative AI for Tutoring: The system could be further developed to include Large Language Models that work in tandem with the SPARQL inference engine to allow learners to ask questions about the course content in real time.

Author Contributions: R.A.S and S.R.M.Z conceptualized, designed the methodology and system. R.A.S was responsible for implementation, software development, data pre-processing, ontology development, execution of the experiments, and formal analysis. S.R.M.Z validation of the proposed framework and overall research guidance. R.A.S drafted the paper and all authors read, commented, revised, edited and took part in preparing the final draft of the manuscript for submission. R.A.S and S.R.M.Z interpreted the results, and agreed on the final content of this paper for submission.

Data Availability Statement: Data used to support the findings of this study were gathered from open access Massive Open Online Course (MOOC) providers (Udemy, Coursera, EDX); and academia courses of the Kurdish language, by using authorized web

scraping and semantic preprocessing approaches. The resultant processed datasets, ontology files, and the implementation tools used in this study are available from the author.

Acknowledgments: Authors would like to thank Akre University for Applied Sciences and Duhok Polytechnic University for providing academic atmosphere, technical equipment, and institutional support that enabled to perform this research. The authors would also like to thank educational platforms and open online learning resources that were used for the data collection and semantic analysis.

Conflicts of Interest: The authors have no conflicts of interests to disclose regarding the publication of this manuscript. The authors state that this work was conducted in the absence of any commercial or financial relationships that could be building as potential conflicts of interest.

Abbreviations: The following abbreviations are used throughout this manuscript:

Abbreviation	Full Term
MOOC	Massive Open Online Course
OWL	Web Ontology Language
RDF	Resource Description Framework
SPARQL	SPARQL Protocol and RDF Query Language
ODHSRS	Ontology-Driven Hybrid Semantic Recommender System
CBF	Content-Based Filtering
SCF	Social Collaborative Filtering
NLP	Natural Language Processing
AI	Artificial Intelligence
ETL	Extract, Transform, Load
XAI	Explainable Artificial Intelligence
LMS	Learning Management System
ICRS	Intelligent Content-Based Recommendation System
MMRS	Multi-Modal Recommendation System

References

1. Agarwal, A.; Mishra, D.S.; Kolekar, S.V. Knowledge-based recommendation system using semantic web rules based on Learning styles for MOOCs. Cogent Eng. 2022, 9, 2022568.
2. Bhutoria, A. Personalized education and Artificial Intelligence in the United States, China, and India: A systematic review using a Human-In-The-Loop model. Comput. Educ. Artif. Intell. 2022, 3, 100068.
3. Mustafa, D.J.; Zeebaree, S.R.M. Optimizing Online Educational Experiences through Semantic Recommender Systems: A Review. Asian J. Res. Comput. Sci. 2025, 18, 211–226.
4. Lan, G.; Liu, T.; Wang, X.; Pan, X.; Huang, Z. A semantic web technology index. Sci. Rep. 2022, 12, 07615-4.
5. Du, W. An Intelligent Agent-Based Collaborative Tutoring System for Mechanical Design Adult Education in Open Universities. In Proceedings of the 2025 International Conference on Educational Technology and Artificial Intelligence; ACM: New York, NY, USA, 2025; pp. 464–469.
6. Wang, Y.; Wang, T.; Wang, Q. The impact of digital transformation on enterprise performance: An empirical analysis based on China's manufacturing export enterprises. PLoS One 2024, 19, e0299723.
7. Duan, C.; Zhang, W.; Cui, Q.; Pei, Y.; He, B.; Huang, Q. Enhancing MOOC Recommendation Through Preference-Aware Knowledge Graph Diffusion and Temporal Sequence Modeling. Information 2025, 16, 1061.
8. Martinez, O.; Sharma, D. Semantic Web Technologies for Knowledge Graph Construction and Querying. Int. J. Adv. Comput. Theory Eng. 2024. Available online: http://creativecommons.org/licenses/by-nc-nd/4.0/ (accessed on Day Month Year).

9. Soto Setzke, D.; Riasanow, T.; Böhm, M.; Krcmar, H. Pathways to Digital Service Innovation: The Role of Digital Transformation Strategies in Established Organizations. Inf. Syst. Front. 2023, 25, 1017–1037.
10. Sazli, W.; Saifudin, N. MPM Model: A Cross-Platform Massive Open Online Course (MOOC) Performance Monitoring and Measurement Model Based on MOOC Learning Analytics. Preprint/Journal Title 2024.
11. Chiang, C.; Lee, H. Recent Advances in Pre-trained Language Models: Why Do They Work and How Do They Work. Conference/Journal Name 2022, 8–15.
12. Hazar, M.J.; Jaballi, S.; Maraoui, M.; Zrigui, M.; Nicolas, H. A Hybrid E-Learning Recommendation System Incorporating User Reviews and Ratings for Enhanced Course Selection. Research Square 2025. doi:10.21203/rs.3.rs-5729775/v1.
13. Wang, L.; Han, W.; Xi, Z. Exploring Semantic Web Tools in Education to Boost Learning and Improve Organizational Efficiency. Int. J. Semant. Web Inf. Syst. 2025, 21.
14. Mustafa, D.J.; Zeebaree, S.R.M. Optimizing Online Educational Experiences through Semantic Ontology-Based Recommender Systems: A Case Study on Coursera. Int. J. Comput. Exp. Sci. Eng. 2025, 11.
15. Ezaldeen, H.; Bisoy, S.K.; Misra, R.; Alatrash, R. Semantics aware intelligent framework for content-based e-learning recommendation. Nat. Lang. Process. J. 2023, 3, 100008.
16. Putra, S.A.; et al. The Development of Intelligent Web for Rural Social E-Learning. J. Nas. Tek. Elektro dan Teknol. Inf. 2024, 13, 112–121.
17. Amin, S.; Uddin, M.I.; Mashwani, W.K.; Alarood, A.A.; Alzahrani, A.; Alzahrani, A.O. Developing a Personalized E-Learning and MOOC Recommender System in IoT-Enabled Smart Education. IEEE Access 2023, 11, 136437–136455.
18. Ma, F. Learning behavior analysis and personalized recommendation system of online education platform based on machine learning. Comput. Educ. Artif. Intell. 2025, 8, 100408.
19. Dutta, S.; Beier, F.; Werth, D. AI-Based Personalized Multilingual Course Recommender System Using Large Language Models. In International Conference on Agents and Artificial Intelligence; SciTePress, 2025; pp. 1069–1076.