IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 31, No. 8, August, 2026, pp. 2579-2590
Journal homepage: http://iieta.org/journals/isi

Ontology-Driven Semantic Information System for Adaptive Assessment in Kazakh-Language Computer Science Education: A Case Study

Rozamgul Niyazova[1], Rakhila Turebayeva[2], Alimbubi Aktayeva[3*], Guldarai Kaparova[3], Zhanyl Sarsenbayeva[3], Kerbez Bolatbekkyzy[4]

[1] Department of Artificial Intelligence Technologies, L.N. Gumilyov Eurasian National University, Astana 01000, Kazakhstan
[2] Department of Information Security, L.N. Gumilyov Eurasian National University, Astana 01000, Kazakhstan
[3] Department of Information Systems and Informatics, A. Myrzakhmetov Kokshetau University, Kokshetau 02000, Kazakhstan
[4] SignBridge LLP, Astana 01000, Kazakhstan

Corresponding Author Email: aakhtaewa@gmail.com

https://doi.org/10.18280/isi.310805

Received: 20 May 2026
Revised: 3 August 2026
Accepted: 14 August 2026
Available online: 31 August 2026

Keywords:
ontology-driven semantic assessment, ontology engineering, semantic similarity, adaptive assessment, Kazakh-language processing, natural language processing, low-resource languages

ABSTRACT

Semantic assessment of learner-generated responses remains challenging in low-resource languages because conventional keyword- and rule-based approaches often fail to capture contextual meaning. This study proposes an ontology-driven semantic information system for adaptive assessment in Kazakh computer science education. The proposed framework integrates ontology engineering, semantic similarity computation, fuzzy semantic matching, and natural language processing (NLP) to transform learner responses into structured semantic representations and evaluate their correspondence with domain knowledge. A computer science ontology was developed for grades 5–11, incorporating domain concepts, semantic relations, hierarchical dependencies, and Kazakh-language terminology. The experimental evaluation was conducted using 300 learner-generated responses, 150 domain-specific question-response pairs, and 1,281 manually validated ontology correspondence pairs. The performance of the system was assessed in terms of accuracy, precision, recall, and F1-score. Experimental results showed that the proposed ontology-driven approach achieved an accuracy of 0.848, precision of 0.85, recall of 0.84, and F1-score of 0.85, outperforming keyword-based and rule-based assessment approaches. The results indicate that ontology-based semantic representation can aid the contextual interpretation of Kazakh-language responses, in addition to direct lexical matching. Although the evaluation is limited to a computer science education case study, the proposed framework provides a structured foundation for developing semantic assessment systems for low-resource language environments.

1. INTRODUCTION

Ontology engineering, semantic web technologies, artificial intelligence, and natural language processing (NLP) support formal knowledge representation, intelligent retrieval, semantic interoperability, and context-aware processing [1, 2]. Ontology-driven information systems structure domain concepts and relations to support semantic interactions in complex digital environments.

However, existing methods often provide limited support for contextual interpretation, adaptive assessment, and explainable processing when information is heterogeneous or linguistically variable. This limitation is particularly evident when systems must interpret user-generated text rather than retrieve predefined content.

Many systems identify correspondences through keyword overlap or lexical similarity. These methods are computationally efficient; however, they may miss conceptually equivalent expressions that share few surface forms, thereby reducing contextual accuracy and interpretability.

This problem is particularly relevant to Kazakh, an agglutinative low-resource language with limited language-specific semantic resources [3-6]. Ontology-based processing can combine linguistic analysis with explicit domain knowledge to address morphological variations and sparse lexical resources.

This study examined the semantic assessment of Kazakh language learner responses in computer science for grades 5–11. The bounded domain provides a controlled setting for testing whether ontology-based representations improve interpretation relative to keyword- and rule-based baselines.

The findings, therefore, apply to the evaluated domain and dataset; they do not establish general effectiveness across all low-resource languages or educational contexts.

Prior works commonly treated ontology representation, semantic similarity, and adaptive assessment as separate functions. Their integration into a single workflow for

assessing Kazakh language learner responses remains insufficiently studied.

To address this gap, this study developed an ontology-driven semantic information system that integrates ontology engineering, linguistic preprocessing, hierarchical similarity evaluation, fuzzy classification, and adaptive recommendation.

The system comprises ontology management, semantic processing, adaptive feedback, and interaction modules. It maps concepts extracted from learner responses to a reference ontology and produces an interpretable assessment based on explicit conceptual relations.

For each response, the system constructs an ontology-based representation, retrieves the corresponding reference structure, calculates the weighted concept-level similarity, and applies fuzzy thresholds to determine the assessment category.

The evaluation used 300 learner-generated responses, 150 domain-specific question–response pairs, and 1,281 expert-annotated ontology correspondence pairs. Performance was measured using accuracy, precision, recall, and F1-score and was compared with keyword- and rule-based baselines.

The main contributions of this study are as follows.

1. An ontology-driven information system architecture for semantic assessment in Kazakh-language computer science education for grades 5–11.

2. Ontology-based representation model for domain concepts, semantic relations, hierarchical dependencies, and contextual mapping.

3. A fuzzy semantic-matching algorithm for comparing ontology structures derived from learner responses with reference knowledge representations.

4. A Kazakh-language NLP pipeline that combines linguistic preprocessing, morphological analysis, syntactic processing, and ontology-based concept extraction.

5. An experimental evaluation was conducted based on learner-generated responses and expert-annotated correspondence pairs using accuracy, precision, recall, and F1-score as the principal metrics.

The remainder of this paper is organised as follows. Section 2 reviews ontology-based information systems, semantic similarity, low-resource NLP, and adaptive semantic architectures. Section 3 describes the ontology, computational model, system architecture, dataset, and evaluation procedure. Section 4 reports and discusses the results, computational implications, scalability, maintenance requirements, and limitations. Section 5 presents the conclusions and future research directions.

2. RELATED WORK

A structured scientometric and literature review was conducted to identify recent research on ontology-driven information systems, semantic technologies, adaptive architectures, and low-resource NLP. The search covered Scopus-indexed publications from 2004 to 2025 and followed a PRISMA-based procedure for identification, screening, eligibility assessment, and inclusion.

The PRISMA procedure provided a transparent and reproducible method for selecting literature relevant to the research problem; it did not form part of the system development or experimental methodology. This review identified the research gap and positioned the proposed semantic assessment system within the existing literature. Figure 1 presents the study selection process.

The search yielded 342 publications. After duplicates were removed and titles, abstracts, and full texts were screened for eligibility, 30 studies were included in the scientometric and methodological analyses. The selected studies were analyzed using scientometric and bibliometric visualization methods to identify thematic trends, the development of the field, and relationships among research topics. Figure 2 shows the principal characteristics of the Scopus dataset.

The analysis indicated increasing research attention on ontology-driven information systems, adaptive semantic architectures, artificial intelligence, and low-resource NLP. It also identified active work in ontology engineering, semantic similarity, intelligent interaction, and adaptive semantic technologies. These findings establish the relevance of ontology-based semantic processing and provide the basis for the methodological and architectural gaps addressed in this study.

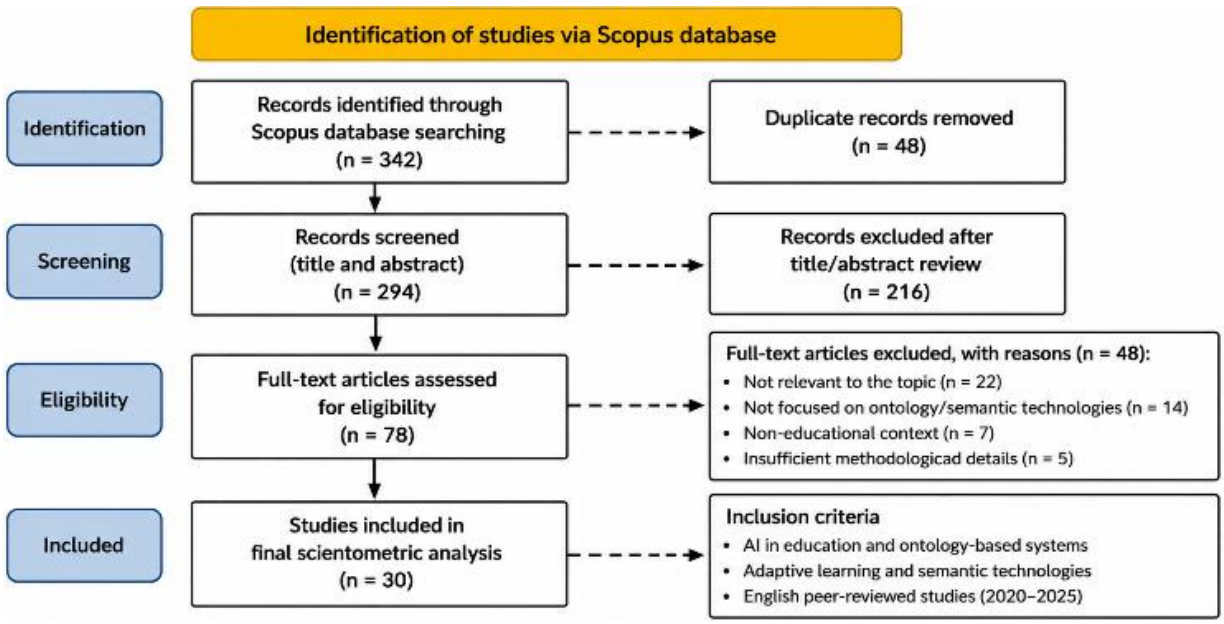

Figure 1. PRISMA-based flow diagram of literature identification and selection for the related-work analysis

Figure 2. Main bibliometric characteristics of the Scopus dataset on ontology-driven semantic information systems

2.1 Ontology-based information systems

Ontology-based information systems support semantic knowledge representation, information integration, intelligent retrieval, and contextual processing in complex digital environments [7-10]. Ontology engineering formally represents domain concepts, semantic relations, hierarchical dependencies, and other types of knowledge.

Established resources, such as the National Cancer Institute Thesaurus and AGROVOC, demonstrate the value of structured semantic vocabularies. Ontology engineering environments, including Computer-aided Ontology Development Architecture (CODA) and Semantic Turkey, support knowledge acquisition and ontology management in diverse domains [11-15].

Previous studies have shown that ontology-driven architectures can improve semantic consistency, contextual interpretation, and interoperability among heterogeneous information resources. Applications include semantic web systems, knowledge-based question answering, educational information management, and adaptive knowledge-processing environments [16-21].

Recent research has emphasized ontology matching and semantic reasoning for intelligent retrieval and automated processing. However, many systems focus on static knowledge representation and organization rather than adaptive interaction or interpretation of user-generated responses. Few studies have examined ontology-driven processing in low-resource languages or multilingual educational systems. Therefore, the integration of low-resource NLP with ontology-based adaptive assessment remains under-explored.

2.2 Semantic similarity and knowledge representation

Semantic similarity analysis supports the contextual interpretation of textual and conceptual information. Keyword matching and lexical similarity can be effective when expressions overlap directly, but they may not recognize conceptually related entities when exact lexical correspondence is absent. Fuzzy semantic analysis and ontology-matching methods allow a more flexible comparison of related information structures [7, 8, 22, 23]. They can also support assessments when lexical correspondence is incomplete.

Ontology- and semantic-graph-based representations explicitly model the conceptual dependencies, hierarchical relationships, and contextual associations. These structures support interpretable semantic retrieval and adaptive information processing. Nevertheless, many semantic evaluation systems provide limited contextual adaptability in multilingual low-resource settings. Ontology-based assessments of Kazakh language responses remain insufficiently studied.

2.3 Natural language processing for low-resource languages

NLP for low-resource languages is an important area of multilingual information systems research. Most advanced models have been developed for high-resource languages, whereas low-resource settings often lack lexical resources, annotated datasets, and language-specific processing tools.

Multilingual methods, cross-lingual transfer learning, and ontology-driven language processing are relevant to low-resource NLP [3-6, 24-26]. However, ontology-based contextual interpretation in agglutinative languages such as Kazakh remains under-explored.

Kazakh presents additional processing challenges owing to its agglutinative morphology, productive suffixation, syntactic variability, and context-dependent interpretation. These properties complicate concept extraction, semantic retrieval, and ontology mapping. Reliable Kazakh-language semantic assessment, therefore, requires morphological analysis, syntactic processing, and domain-based semantic interpretation before ontology mapping.

2.4 Adaptive semantic information systems

Adaptive semantic information systems integrate ontology-based processing, recommendation mechanisms, contextual interaction models, and knowledge management (KM). These components can improve semantic interoperability, feedback generation, and adaptive interactions [17-21, 27-32].

Previous studies have reported improvements in contextual retrieval, semantic consistency, personalized support, and interaction management. Ontology-driven systems also make conceptual dependencies and assessment decisions traceable.

However, systems that rely primarily on keyword retrieval or fixed answer rules still provide limited contextual interpretations.

The proposed system integrates functions that prior approaches have often treated separately. Kazakh-language responses were linguistically pre-processed, mapped to concepts and relations in a domain ontology, converted into ontology-based response structures, and compared with reference structures using weighted similarity and fuzzy thresholds. The resulting assessment provided adaptive feedback and recommendations.

Ontology mapping was, therefore, an intermediate operation rather than the final objective. The contribution lies in combining ontology mapping, concept-level comparison, fuzzy classification, and adaptive feedback within a domain-specific assessment workflow that does not replace general-purpose ontology-alignment methods.

3. MATERIALS AND METHODS

3.1 Research design and system development procedure

Adaptive semantic information systems integrate ontology-based processing, recommendation mechanisms, contextual interaction models, and knowledge management (KM). These components can improve semantic interoperability, feedback generation, and adaptive interactions [17-21, 27-32].

Previous studies have reported improvements in contextual retrieval, semantic consistency, personalized support, and interaction management. Ontology-driven systems also make conceptual dependencies and assessment decisions traceable. However, systems that rely primarily on keyword retrieval or fixed answer rules still provide limited contextual interpretations.

The proposed system integrates functions that prior approaches have often treated separately. Kazakh-language responses were linguistically pre-processed, mapped to concepts and relations in a domain ontology, converted into ontology-based response structures, and compared with reference structures using weighted similarity and fuzzy thresholds. The resulting assessment provided adaptive feedback and recommendations.

Ontology mapping was, therefore, an intermediate operation rather than the final objective. The contribution lies in combining ontology mapping, concept-level comparison, fuzzy classification, and adaptive feedback within a domain-specific assessment workflow that does not replace general-purpose ontology-alignment methods.

3.2 Input and output data specification

The system accepted domain-specific questions, learner responses, ontology concepts, semantic relations, and reference knowledge as inputs. It mapped extracted response concepts to the ontology, compared the resulting representation with reference knowledge, and generated an assessment label, feedback, and a recommendation. The inputs were defined as follows:

$Q = \{q_1, q_2, ..., q_m\}$: set of domain-specific semantic questions from the computer science domain.

$R = \{r_1, r_2, ..., r_k\}$: set of learner-generated responses.

$O = \{C, E, H, W\}$: domain ontology, where C is the set of concepts, E is the set of semantic relations, H is the set of hierarchical dependencies, and W is the set of concept weights.

$K = \{k_1, k_2, ..., k_n\}$: set of ontology-based reference knowledge structures stored in the knowledge base.

$L = \{\ell_1, \ell_2, \ldots, \ell_p\}$: set of linguistic units obtained by preprocessing and analyzing Kazakh-language learner responses.

The output data are defined as follows:

$S(O_r, O_k)$: semantic similarity score between the response ontology O_r and the reference ontology O_k.

$y \in$ {correct, partially correct, incorrect}: semantic assessment label.

F: adaptive feedback generated from the semantic-matching result.

A: adaptive recommendation for supplementary content or knowledge revision.

These definitions specify the complete input–processing–output sequence used by the proposed system.

3.3 Ontology construction

The ontology was developed as a formal representation of computer science knowledge for grades 5–11. Protégé was used to construct and visualize the ontology and verify its logical consistency. Figure 3 shows a fragment of the resulting structure and its semantic relations.

Figure 3 shows the links among grade-level concepts, prerequisite structures, glossary entries, translations, and knowledge units in the computer science curriculum. The ontology construction comprises six steps. Key concepts were extracted from curriculum materials and thesaurus resources, semantic relations were identified, and hierarchical structures were created. Concept weights were then assigned, semantic consistency was verified, and ontology modules were integrated into the information system.

The ontology contains domain concepts, semantic relations, hierarchical dependencies, prerequisite links, and Kazakh language lexical units. The principal relation types are "is-a", "part-of", "related-to", and "prerequisite-of".

The knowledge base comprises grade-specific Web Ontology Language (OWL) modules for computer science curricula in grades 5–11. The modules represent domain concepts, curriculum entities, glossary and multilingual lexical resources, and semantic relationships. Raw OWL class declarations were distinguished from the normalized class graph used for structural evaluations.

For structural analysis, grade-specific modules were merged into a directed OWL class graph. The node set V contained unique named OWL classes after identifiers were normalized and identical classes across modules were consolidated.

The edge set E contained explicit class-to-class *rdfs:subClassOfrelations* and non-lexical object-property relations between named classes. Assertions using the '*Дегеніміз*' property were excluded because they encode lexical or multilingual correspondence rather than structural concept-to-concept relations. Duplicate nodes and directed edges were counted once. The resulting ontology serves as the semantic knowledge base of the system and supports response assessment, question answering, recommendation generation, and contextual interpretation.

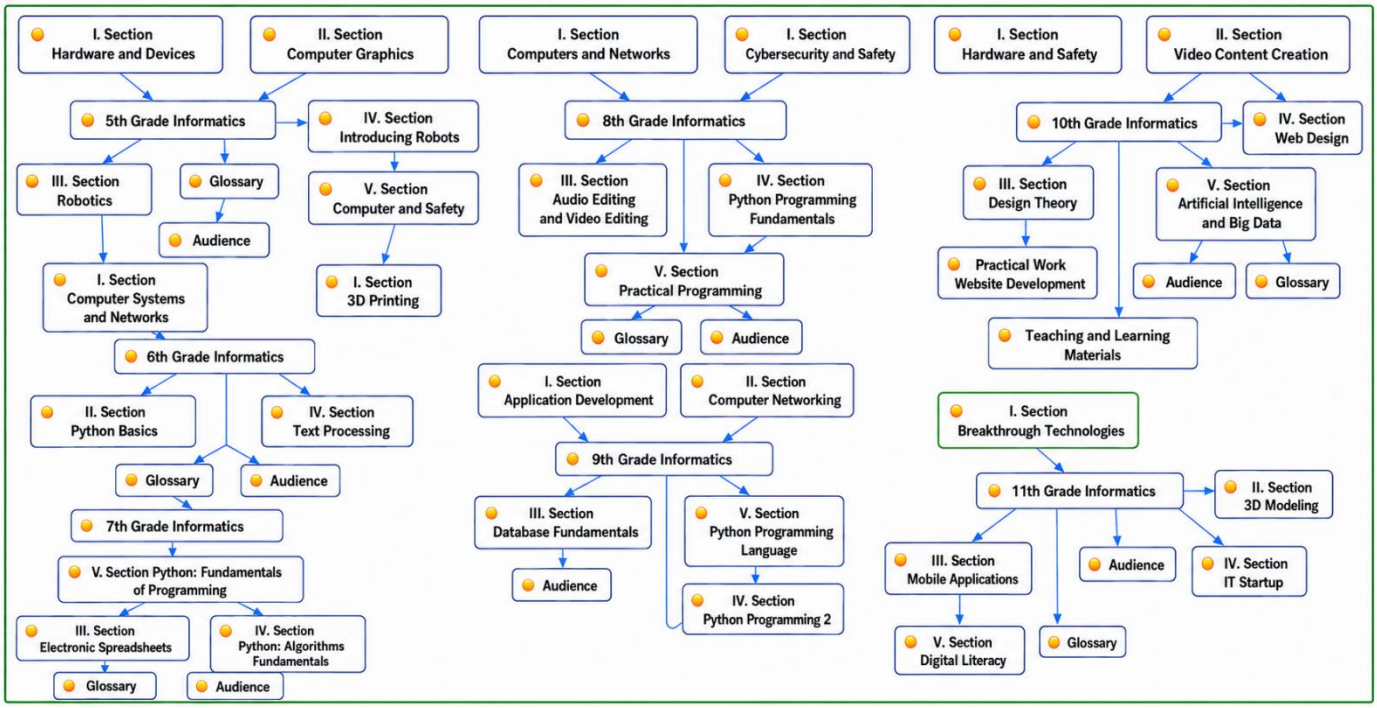

Figure 3. Fragment of the developed computer science ontology for grades 5–11

3.4 Architecture of the ontology-driven semantic information system

The system was implemented as a multilayer architecture with six interconnected layers, as shown in Figure 4:

1. *Client interaction layer:* This layer provides web, mobile, and API interfaces for content access, query submission, response entry, report viewing, and user interaction.

2. *Semantic processing platform:* This layer provides storage, computation management, security services, and API-based integration for the distributed components.

3. *Semantic processing layer:* This layer performs linguistic preprocessing, semantic parsing, ontology mapping, similarity calculation, ontology matching, and fuzzy evaluation. It transforms recognized linguistic units into an ontology-based response structure aligned with domain concepts, semantic relations, and reference knowledge.

4. *Adaptive semantic layer:* This layer analyses assessment results and interaction history to generate feedback, recommendations, and context-aware learner support.

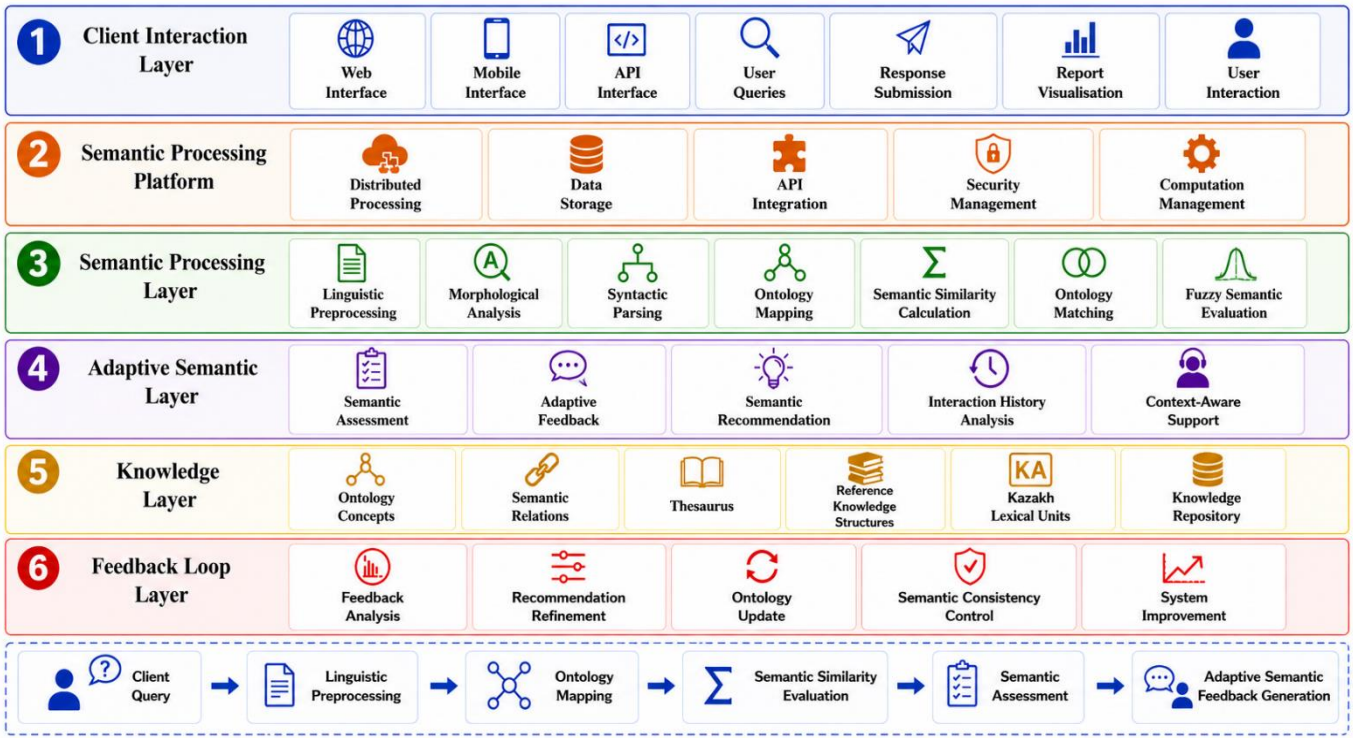

Figure 4. Architecture of the proposed ontology-driven semantic information system

5. *Knowledge layer:* This layer stores the ontology, semantic relations, lexical resources, reference structures, and domain content required for processing.

6. *Feedback loop layer:* This layer supports feedback analysis, recommendation refinement, ontology updates, semantic-consistency control, and system improvement.

The layers operate sequentially: the system receives a learner response, performs linguistic preprocessing and ontology mapping, calculates semantic similarity, assigns an assessment label, and generates adaptive feedback.

This architecture supports contextual interpretation and explicit identification of semantic correspondence beyond direct lexical matching with predefined answer templates.

Figure 4 shows the functional architecture implemented for the case study, not a production-scale deployment. This study evaluated the semantic-processing workflow and assessment performance using an experimental dataset. It did not benchmark end-to-end latency, throughput, concurrent user capacity, resource utilization, or network performance under operational workloads.

Therefore, the reported results demonstrate functional and semantic performance within the prototype environment. They do not establish production-scale runtime performance, which requires separate deployment-oriented evaluations under realistic workloads.

3.5 Semantic similarity evaluation model

The computational model was implemented as a sequential pipeline that transformed each learner response into an ontology-based representation and generated assessment and adaptive feedback. The processing comprised linguistic preprocessing, concept extraction, ontology mapping, construction of the response ontology, retrieval of the reference ontology, concept-level similarity calculation, weighted aggregation, fuzzy classification, and feedback generation.

The model mapped an input response r, together with the domain ontology O and reference knowledge structure K, to the output tuple (S, y, F, A), where S is the aggregated semantic similarity score, y is the semantic assessment label, F is adaptive feedback, and A is the adaptive recommendation.

Semantic similarity measured the correspondence between the ontology structure derived from a learner response and the relevant reference structure in the knowledge base.

Let O_r denote the ontology derived from the learner response and O_k the reference ontology. The aggregated semantic similarity score was calculated using Eq. (1):

$$(O_r, O_k) = \frac{\sum_{i=1}^{n} \omega_i \cdot sim(c_{ri}, c_{ki})}{\sum_{i=1}^{n} \omega_i} \quad (1)$$

where,

$S(O_r, O_k)$ is the aggregated semantic similarity score;

O_r is the ontology structure generated from the learner response;

O_k is the reference ontology structure from the knowledge base;

c_{ri} is the i-th concept extracted from the learner response;

c_{ki} is the corresponding reference concept;

$\text{sim}(c_{ri}, c_{ki})$ is the semantic similarity between the two concepts; and ω_i is the manually assigned weight of the i-th concept.

The weights were specified as domain-dependent parameters and were not estimated from the experimental dataset.

Concept-level similarity $\text{sim}(c_{ri}, c_{ki})$ was defined as a depth-based hierarchical measure over the *rdfs:subClassOf* structure:

$$sim(c_{ri}, c_{ki}) = \frac{2\ depth(LCS(c_{ri}, c_{ki}))}{depth(c_{ri}) + depth(c_{ki})} \quad (2)$$

where, $LCS(c_{ri}, c_{ki})$ denotes the lowest common subsumer of the two concepts, and depth (c) denotes the distance from concept c to the relevant ontology root.

Similarity ranges from 0 to 1; a value of 1 indicates identical concepts, and higher values indicate closer structural proximity. Synonyms, hypernyms, hyponyms, meronyms, and holonyms in the domain thesaurus supported the preceding concept-mapping stage but were not included as separate weighted components in Eq. (2).

The concept weights ω_i were assigned during ontology construction according to the importance and hierarchical position of each concept. They were neither learned from the response dataset nor optimised on an independent validation set. Therefore, the weighting scheme is a domain-specific ontology parameterisation rather than a universally calibrated model. Here, n denotes the number of concept pairs compared.

The aggregated score was interpreted using the following fuzzy thresholds:

- High semantic correspondence: $S \geq 0.80$.
- Partial semantic correspondence: $0.50 \leq S < 0.80$.
- Low semantic correspondence: $S < 0.50$.

The thresholds of 0.50 and 0.80 were predefined as operational boundaries for low, partial, and high correspondence in the evaluated computer science domain. They are not assumed to be universally valid and require recalibration with domain-specific validation data before transfer to other subjects or ontology structures.

The model can therefore recognise related ontology concepts even when a learner's response does not reproduce the reference wording.

3.6 Semantic matching algorithm

Algorithm 1 specifies the comparison of a response-derived ontology structure with the corresponding reference structure and the generation of assessment outputs.

Algorithm 1. Ontology-based semantic matching algorithm

Input: learner response r, semantic question q, domain ontology O, and reference knowledge structure K.
Output: similarity score S, assessment label y, adaptive feedback F, and recommendation A.
Step 1. Receive the learner response r.
Step 2. Apply tokenisation and morphological analysis to r.
Step 3. Extract candidate domain concepts from the processed response.
Step 4. Map the extracted concepts to the ontology concepts in O.
Step 5. Construct the response ontology O_r.
Step 6. Retrieve the reference ontology O_k from K.
Step 7. For each concept pair (c_{ri}, c_{ki}), calculate the depth-based similarity defined in Eq. (2).
Step 8. Assign the manually defined weights ω_i according to concept importance and hierarchical position.
Step 9. Calculate $S(O_r, O_k)$ using Eq. (1).
Step 10. Apply the fuzzy thresholds to classify the response.
Step 11. Generate the assessment label y.
Step 12. Generate adaptive feedback F and recommendation A.
Step 13. Return S, y, F, and A.

The algorithm supports contextual assessment by comparing concepts and ontology relations rather than isolated keywords.

3.7 Kazakh-language natural language processing module

The Kazakh language processing module was developed to interpret learner responses in a low-resource setting. Because Kazakh expresses grammatical and lexical information through productive suffixation and word formation, direct keyword matching does not provide a reliable basis for semantic assessment. Therefore, linguistic preprocessing was performed before ontology mapping and similarity evaluation.

The NLP module linked language-specific analysis to the assessment workflow through tokenization, morphological analysis, and the extraction of candidate domain concepts.

3.8 Experimental dataset

The experimental resources comprised an ontology and response-based evaluation dataset. Grade-specific OWL modules for computer science in grades 5–11 were converted into a merged graph using the procedure described in Section 3.3. After class identifiers were normalised, identical classes were consolidated, and duplicate directed edges were removed. The graph contained 1,280 unique OWL classes and 1,339 unique directed structural edges.

The evaluation dataset contained 300 learner-generated responses and 150 domain-specific question–response pairs. Mapping the concepts extracted from the responses to their reference ontology concepts produced 1,281 correspondence pairs for quantitative evaluation.

Three domain experts in ontology engineering, semantic technologies, and intelligent information systems manually assigned reference labels to 1,281 correspondence pairs. The original protocol did not quantify inter-annotator agreement, and individual disagreement records were not retained. Therefore, agreement coefficients and disagreement frequencies could not be calculated retrospectively; this limitation should be considered when interpreting the results.

The dataset served as a fixed case study corpus rather than a training sample for a statistical model. Concept weights were specified manually, and fuzzy thresholds were predefined; no parameters were learned from the responses. Training–test splitting and k-fold cross-validation were therefore not applicable to the implemented rule- and ontology-based procedures. However, the modest corpus size limits the empirical scope of the findings, which should be interpreted as case-study evidence rather than population-level estimates of generalisation.

A fixed reference set was used to assess the correspondence classification, contextual interpretation, and consistency of the adaptive output. Future studies should use larger independent datasets from additional subjects, institutions, age groups, and linguistic settings.

3.9 Evaluation metrics

The system was evaluated using standard classification metrics for binary ontology-correspondence decisions. A true positive (TP) was a correctly identified correspondence, whereas a true negative (TN) was a correctly rejected non-correspondence. A false positive (FP) was an incorrect match, and a false negative (FN) was a relevant correspondence that the system failed to identify.

Accuracy, precision, recall, and F1-score were calculated from these outcomes. The proposed system was compared with two baseline methods:

1. Keyword-based assessment based on direct lexical overlap.
2. Rule-based assessment based on predefined answer patterns.

These lightweight baselines were part of the original experimental design. Embedding-based semantic similarity models were not included. All three methods were evaluated using the same learner responses and expert-annotated correspondence pairs.

4. RESULTS AND DISCUSSION

4.1 Ontology structure and graph statistics

The developed ontology comprises grade-specific OWL modules for computer science in grades 5–11. It represented domain concepts, semantic relations, hierarchical and prerequisite structures, glossary resources, and multilingual lexical units derived from curriculum materials. For structural evaluation, these modules were transformed into the normalized merged graph defined in Section 3.3.

After class identifiers were normalized and identical classes across grade-specific modules were consolidated, the graph contained 1,280 unique named OWL class nodes and 1,339 unique directed structural edges. Its density was approximately 0.0818%. The mean in-degree and out-degree were each approximately 1.046, giving a mean total degree of approximately 2.092. These values characterize the implemented ontology as a sparse, predominantly hierarchical graph.

This structure provides the domain representation used to interpret learner responses in the proposed intelligent information system.

4.2 Semantic matching performance

The semantic assessment mechanism was evaluated using 300 learner-generated responses and 1,281 manually annotated correspondence pairs. The analysis considered correspondence with the reference ontology, mapping consistency, contextual interpretation, and stability of the assessment output. Table 1 summarises the four correspondence outcomes used to calculate the classification metrics.

Table 1. Ontology correspondence classification results

Classification Result	Value
True Positives (TP)	531
False Positives (FP)	94
False Negatives (FN)	101
True Negatives (TN)	555
Total evaluated correspondences	1281

Table 1 shows 531 TPs, 555 TNs, 94 FPs, and 101 FNs. The resulting accuracy was (531 + 555) / 1,281 = 0.8478, rounded to 0.848. The evaluation metrics were calculated as follows:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} = \frac{(531 + 555)}{1281} \approx 0.85$$

$$Precision = TP/(TP + FP) = 531/(531 + 94) = 0.85$$

$$Recall = TP/(TP + FN) = 531/(531 + 101) = 0.84$$

$$F1 - score = 2 \times (Precision \times Recall)/(Precision + Recall) = 0.85$$

The 94 FPs and 101 FNs indicated limitations at several stages of the pipeline. FPs may occur when individually relevant concepts produce a high aggregated score, even if the complete relational meaning does not match the reference. FNs may occur when paraphrases, lexical variants, or Kazakh-language constructions are not captured during concept extraction or ontological mapping. Manually assigned weights and fixed fuzzy thresholds may also affect borderline cases.

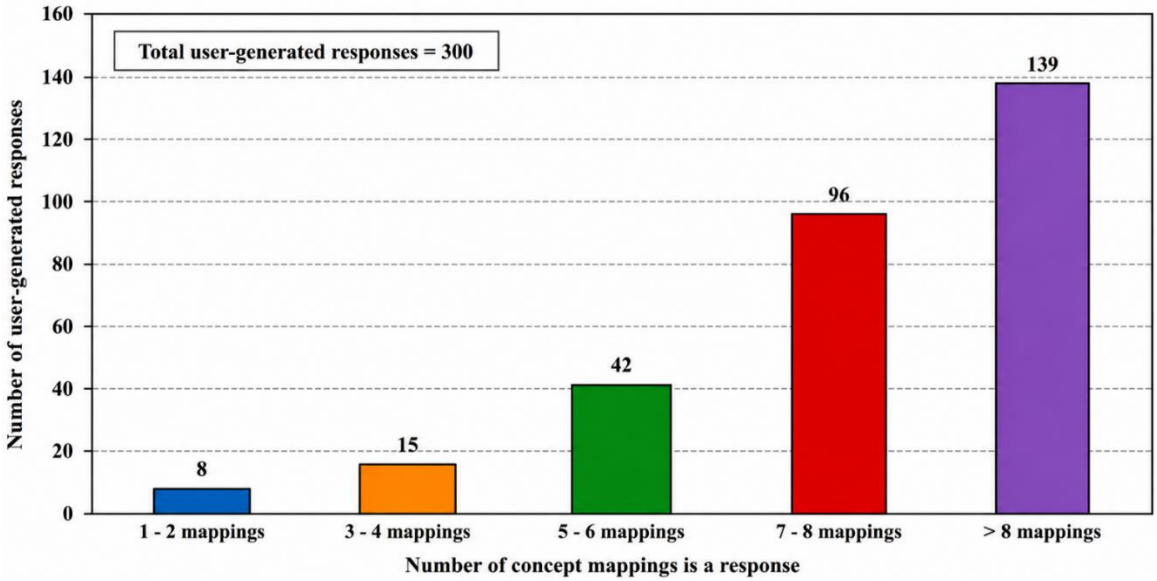

Figure 5. Distribution of user-generated responses by the number of ontology concept mappings per response

These observations identified concept extraction, ontology coverage, relational representation, and decision thresholds as plausible error sources. Because the original evaluation did not use a predefined error taxonomy or retain category labels for individual errors, their frequencies could not be quantified retrospectively. Figure 5 presents the distribution of the mapped ontology concepts across the 300 responses. Overall, the system identifies contextual correspondences that are not limited to direct lexical overlaps.

4.3 Confidence interval analysis

A 95% confidence interval was calculated for semantic-matching accuracy using the normal approximation (Wald interval) for a binomial proportion.

With 1,086 correctly classified ontology correspondence pairs out of 1,281 observations $\hat{\rho} = 0.8478$, the interval was calculated as $\hat{\rho} \pm \frac{1.96\sqrt{\hat{\rho}(1-\hat{\rho})}}{n}$, to yield a 95% confidence interval of approximately [0.83, 0.87].

The resulting interval indicates the statistical uncertainty associated with the estimated semantic-matching accuracy within the evaluated dataset.

4.4 Example of ontology-based semantic matching

The following example illustrates an ontology-based assessment procedure: The semantic question was: '*What is information coding?*'

The learner response was: "*Coding means transforming information into symbols for storage and transmission*."

After Kazakh-language preprocessing and ontology mapping, the system extracts the concepts of information, coding, transformation, storage, and transmission. The reference structure contains information, coding, representation, storage, and communication.

The system matched '*transformation*' with '*representation*' and '*transmission*' with '*communication*'. The resulting score was $S(O_r, O_k) = 0.89$, and the responses were classified as follows.

1. Semantic assessment label: *Correct.*
2. Semantic correspondence level: *High.*

This example demonstrates how ontology relations support correct assessment when a learner response differs lexically from the reference representation.

4.5 Comparative evaluation

The proposed system was compared with keyword-based lexical matching and rule-based assessment based on predefined answer patterns. All methods were evaluated using the same response and correspondence pairs.

Table 2 shows that the ontology-driven system achieved higher accuracy, precision, recall, and F1-score than both evaluated baselines. This comparison was limited to the selected lightweight baselines and did not establish superiority over the embedding-based methods.

Table 2. Comparative evaluation of assessment approaches

Evaluation Method	Semantic Accuracy	Precision	Recall	F1-Score
Keyword-based matching	0.69	0.67	0.65	0.66
Rule-based assessment	0.72	0.71	0.69	0.70
Proposed ontology-based system	0.848	0.85	0.84	0.85

The largest improvement occurred for conceptually related responses without direct lexical overlap, which is consistent with the intended function of ontology-based matching.

4.6 Adaptive semantic interaction results

The adaptive module generates feedback and recommendations based on the assessment results and the learner's interaction history. Recommendations targeted ontology concepts that were absent or weakly represented in the responses.

Three domain experts assessed the recommendation relevance using three criteria: correspondence with the identified weak or missing concept, consistency with the assessment result, and usefulness in directing the learner to the appropriate reference knowledge. A recommendation was labelled relevant when it met at least two criteria. Table 3 summarizes the results.

Table 3. Adaptive recommendation performance

Metric	Value
Adaptive recommendations generated	300
Relevant recommendations	256
Recommendation accuracy	0.85

The experts classified 256 of the 300 recommendations as relevant with an accuracy of 0.85. Thus, most generated recommendations were consistent with expert judgements under the stated criteria. These findings indicate that ontology-based assessment can support context-aware recommendations and personalized feedback within the evaluated domain.

4.7 System effectiveness and interpretation

The experimental results suggest that ontology-driven semantic technologies can improve contextual semantic interactions and the interpretation of user-generated responses within intelligent semantic information processing systems. Within the evaluated prototype, the system

1. Performed ontology-based semantic assessment.
2. Identified conceptual relations between learners and reference representations.
3. Generated adaptive feedback and recommendations.
4. Processed Kazakh language learner responses.
5. Provided context-aware support in a low-resource language setting.

The higher performance relative to the selected baselines indicates that explicit ontology relations contributed information beyond the keyword overlap. The following subsections interpret these findings in relation to computational complexity, scalability, ontology maintenance, methodological limitations, and alternative approaches to semantic processing.

4.8 Computational implications of ontology-driven semantic processing

The results indicate that explicit ontology structures can improve correspondence identification when learners and reference responses use different lexical forms. This interpretation is restricted to the evaluated dataset for computer science in Kazakhs.

This architecture separates linguistic pre-processing, ontology mapping, similarity calculation, classification, and feedback generation. This modular structure makes each processing decision traceable to recognized concepts, relationships, weights, and thresholds.

The ontology also explicitly represents hierarchical and prerequisite relations, which supports interpretable feedback and controlled knowledge management. No runtime benchmark was conducted; the findings concern semantic and functional performance rather than computational efficiency.

The case study provides a basis for testing comparable symbolic or hybrid architectures in other low-resource language and educational settings.

4.9 System scalability and architectural considerations

The multilayer architecture separates client interaction, linguistic and semantic processing, recommendation, and ontology-based storage. This modular organization permits individual components to be extended or scaled independently.

Separating preprocessing, ontology mapping, evaluation, and knowledge management also simplifies maintenance and integration with different computing environments.

Because the ontology and knowledge base are independent of the client interface, the architecture can support different devices, external semantic platforms, and cloud-based services.

Nevertheless, scalability depends on the ontology size, relation density, and interaction volume. An increase in these factors may increase the processing latency, retrieval complexity, and graph-traversal overhead.

Semantic indexing, consistency validation, and repeated similarity calculations over evolving ontologies may incur additional computational costs.

Let n be the number of concepts extracted from a response and m the number of candidate reference concepts. Direct pairwise comparison has the worst-case complexity $O(n, m)$, whereas weighted aggregation is linear in the number of evaluated concept pairs. The actual cost depends on indexing, candidate selection, relation density, and reference-set size; this bound does not represent the measured end-to-end performance.

The runtime performance was not benchmarked because the experiment focused on correspondence identification. End-to-end latency, throughput, concurrent user capacity, processor and memory utilization, and network overhead were therefore not measured. The implementation should be regarded as a functional research prototype rather than a production-ready deployment.

Future deployment studies should measure the response time, ontology-mapping latency, matching time, throughput, resource utilization, and performance as ontology size and concurrent loads increase. Graph databases, semantic indexes, ontology caching, and parallel similarity calculations should also be evaluated as optimization strategies.

4.10 Ontology maintenance and knowledge base evolution

Ontology maintenance is a central limitation of ontology-driven systems. Domain models must be updated when curricula, terminology, knowledge, or prerequisite relationships change.

The prototype did not implement ontology versioning or lifecycle management functions. Construction and refinement were performed manually without automated change tracking,

version comparison, rollback, or lifecycle states. Therefore, maintenance and controlled evolution are requirements for future deployment rather than demonstrated capabilities.

As ontology grows, preserving semantic integrity and preventing inconsistent mappings will require increasingly systematic validation.

Extending the ontology beyond computer science or the grades 5–11 curriculum would require new concept modelling, relation verification, multilingual adaptation, and domain-specific knowledge engineering.

Incomplete mapping, inconsistent hierarchies, and misaligned dependencies may reduce the correspondence accuracy and quality of adaptive feedback.

Semi-automated refinement and alignment may reduce the burden on domain experts, whereas automated consistency checks can support reliable knowledge base evolution.

Future implementations should record explicit version identifiers and change logs, validate the consistency after each update, and provide controlled comparisons and rollback between ontology versions.

4.11 Limitations of rule-based and fuzzy semantic approaches

Several methodological limitations qualify the reported performance of the rule- and ontology-based assessment procedures.

The evaluation used 300 learner responses and 150 question–response pairs from one educational domain. Although these data produced 1,281 expert-annotated correspondence pairs, the corpus is too limited to establish generalization across subjects, learner populations, age groups, institutions, or languages. Because the method did not learn parameters from the corpus, conventional training–test splitting and k-fold cross-validation were not applied. The findings should therefore be interpreted as case study evidence of feasibility.

Fuzzy assessment depends on manually specified concept weights and predefined thresholds. Therefore, the performance may vary with ontology design, weighting decisions, and calibration procedures.

No sensitivity analysis examined alternative values for the 0.50 and 0.80 thresholds. Therefore, the robustness of the classifications to these boundaries is unknown. Before transferring to another domain, the thresholds should be recalibrated using independent domain-specific validation data.

Symbolic ontology representations provide interpretable relations and transparent decision paths; however, they may represent ambiguous or implicit meanings less effectively. Ontology construction, relation definition, and validation require domain expertise, and the associated maintenance effort increases with system size and complexity.

This method does not use probabilistic representations or neural embeddings. Therefore, its performance depends on the preprocessing accuracy, ontology completeness, and relation coverage. This design favors traceability but may reduce flexibility in dynamic or linguistically ambiguous settings.

Hybrid models that combine ontology engineering with semantic embedding and transformer-based processing should be investigated to improve adaptability while retaining interpretability.

Embedding-based baselines were not included in the quantitative evaluation. Therefore, the results do not establish whether the ontology-driven system outperforms multilingual contextual or subword-based models. Future studies should use the same responses, reference labels, and evaluation metrics.

The manually defined concept weights were not optimized on an independent validation set. Therefore, subjectivity in the weighting scheme cannot be excluded. Sensitivity analysis, structured expert consensus, and data-driven weighting methods should be evaluated.

Inter-annotator agreement was not quantified, and individual expert labels were discarded. Future studies should use a predefined annotation protocol, preserve individual judgements, report agreement coefficients, and document how disagreements are resolved.

4.12 Comparison with knowledge graph, embedding, and large language models-based approaches

This subsection provides a conceptual comparison rather than an additional quantitative benchmark. Embedding-based and large language model (LLM) baselines were excluded from the experiment, so this study makes no claims of empirical superiority over these approaches.

Knowledge graphs, semantic embeddings, transformer architectures, and LLMs are increasingly being used for semantic retrieval, contextual processing, and educational interactions. Recent reviews document both the opportunities and limitations of LLMs in education [33-35].

Each approach differs in terms of requirements for data, computation, interpretability, and domain control. These differences provide the basis for positioning the proposed symbolic method within current semantic information system research.

Compared with embedding-based representations, ontologies express symbolic relations directly. The proposed method records the domain hierarchies, prerequisites, and correspondence decisions explicitly, which supports traceable assessment.

Knowledge graphs provide flexible representations and large-scale retrievals across heterogeneous resources. However, their construction may require extensive linked data, complex integration procedures, and specialized graph-processing infrastructures.

Transformer-based methods can model contextual language variation without an explicit ontology; however, they commonly require substantial pre-trained resources and computational capacity. In contrast, the proposed system applies domain-controlled mapping and thresholds in a resource-constrained setting.

LLMs support contextual interpretation and natural-language generation, but their probabilistic outputs may be difficult to trace to specific domain relationships. Their data and infrastructure requirements may also limit their deployment in low-resource language settings.

The proposed architecture prioritizes explicit domain representation, traceability, and controlled correspondence decisions. These properties suit the assessment tasks in which the basis of a decision must be inspected.

Future work should compare ontology-, embedding-, and LLM-based methods on the same Kazakh-language dataset. Hybrid architectures should also be evaluated to determine whether contextual flexibility can be improved without sacrificing traceability. These comparative statements remain conceptual until the controlled experiments are completed.

5. CONCLUSIONS

This study developed and evaluated an ontology-driven information system for adaptive semantic assessment in Kazakh-language computer science education for grades 5–11. The architecture combines ontology engineering, linguistic preprocessing, hierarchical similarity, fuzzy classification, and adaptive recommendation.

The system interprets learner responses by mapping the extracted concepts to a reference ontology and evaluating their weighted structural similarity. This explicit representation makes the assessment process more traceable than direct keyword or fixed-rule matching.

Within the evaluated case study, the proposed system achieved stronger correspondence-classification performance than both selected baselines and generated recommendations that were usually consistent with expert judgements. The findings support the feasibility of ontology-based contextual assessment in the investigated Kazakh language domain.

The conclusions were limited by the single-subject domain, modest fixed dataset, manually constructed ontology, expert-defined weights and thresholds, absence of embedding-based baselines, and lack of formal inter-annotator agreement and runtime benchmarking.

Future research should validate the framework on larger independent datasets and additional subjects, systematically calibrate weights and thresholds, and compare it with multilingual embedding and LLM-based methods. Deployment studies should also evaluate scalability, resource use, ontology lifecycle management, and hybrid symbolic–neural architectures.

REFERENCE

[1] Abu-Salih, B., Alotaibi, S. (2024). A systematic literature review of knowledge graph construction and application in education. Heliyon, 10(3): e25383. https://doi.org/10.1016/j.heliyon.2024.e25383

[2] Shimizu, C., Hitzler, P. (2025). Accelerating knowledge graph and ontology engineering with large language models. Journal of Web Semantics, 85: 100862. https://doi.org/10.1016/j.websem.2025.100862

[3] Yazar, B.K., Kiliç, E. (2025). Improving low-resource Kazakh-English and Turkish-English neural machine translation using transfer learning and part of speech tags. IEEE Access, 13: 32341-32356. https://doi.org/10.1109/access.2025.3542491

[4] Togmanov, M., Mukhituly, N., Turmakhan, D., et al. (2025). KazMMLU: Evaluating language models on Kazakh, Russian, and regional knowledge of Kazakhstan. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vienna, Austria, pp. 14403-14416. https://doi.org/10.18653/v1/2025.acl-long.701

[5] Kadyrbek, N., Tuimebayev, Z., Mansurova, M., Viegas, V. (2025). The development of small-scale language models for low-resource languages, with a focus on Kazakh and direct preference optimization. Big Data and Cognitive Computing, 9(5): 137. https://doi.org/10.3390/bdcc9050137

[6] Tleubayeva, A., Aubakirov, S., Tabuldin, A., Shomanov, A. (2025). Development and evaluation of a small Kazakh language corpus to improve the efficiency of multilingual NLP systems in low-resource environments. In 2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan, pp. 1-6. https://doi.org/10.1109/SIST61657.2025.11139363

[7] Kollapally, N.M., Geller, J., Keloth, V.K., He, Z., Xu, J. (2025). Ontology enrichment using a large language model: Applying lexical, semantic, and knowledge network-based similarity for concept placement. Journal of Biomedical Informatics, 168: 104865. https://doi.org/10.1016/j.jbi.2025.104865

[8] Zhang, F., Yang, P., Li, R., et al. (2024). A multi-strategy ontology mapping method based on cost-sensitive SVM. Journal of Cloud Computing, 13: 144. https://doi.org/10.1186/s13677-024-00708-7

[9] Wang, L., Sun, C., Zhang, C., Nie, W., Huang, K. (2023). Application of knowledge graph in software engineering field: A systematic literature review. Information and Software Technology, 164: 107327. https://doi.org/10.1016/j.infsof.2023.107327

[10] Tang, Z., Li, T., Wu, D., Liu, J., Yang, Z. (2024). A systematic literature review of reinforcement learning-based knowledge graph research. Expert Systems with Applications, 238: 121880. https://doi.org/10.1016/j.eswa.2023.121880

[11] Sioutos, N., Coronado, S. de, Haber, M.W., Hartel, F.W., Shaiu, W.L., Wright, L.W. (2007). NCI Thesaurus: A semantic model integrating cancer-related clinical and molecular information. Journal of Biomedical Informatics, 40(1): 30-43. https://doi.org/10.1016/j.jbi.2006.02.013

[12] Caracciolo, C., Stellato, A., Morshed, A., et al. (2013). The AGROVOC linked dataset. Semantic Web, 4(3): 341-348. https://doi.org/10.3233/sw-130106

[13] Fiorelli, M., Gambella, R., Pazienza, M.T., Stellato, A., Turbati, A. (2014). Semi-automatic knowledge acquisition through CODA. In Lecture Notes in Computer Science, Kaohsiung, Taiwan, pp. 78-87. https://doi.org/10.1007/978-3-319-07467-2_9

[14] Fiorelli, M., Pazienza, M.T., Stellato, A., Turbati, A. (2014). CODA: Computer-aided ontology development architecture. IBM Journal of Research and Development, 58(2/3): 14:1-14:12. https://doi.org/10.1147/jrd.2014.2307518

[15] Pazienza, M.T., Scarpato, N., Stellato, A., Turbati, A. (2012). Semantic Turkey: A browser-integrated environment for knowledge acquisition and management. Semantic Web, 3(3): 279-292. https://doi.org/10.3233/sw-2011-0033

[16] Diefenbach, D., Lopez, V., Singh, K., Maret, P. (2017). Core techniques of question answering systems over knowledge bases: A survey. Knowledge and Information Systems, 55(3): 529-569. https://doi.org/10.1007/s10115-017-1100-y

[17] Sassi, N., Jaziri, W. (2026). Integrating ontology and knowledge graphs for intelligent assessment and feedback in e-learning systems. Scientific Reports, 16: 24869. https://doi.org/10.1038/s41598-026-54449-5

[18] Csépányi-Fürjes, L., Kovács, L. (2026). Intelligent tutoring in dynamic domains: A graph-based system for comparative analysis of adaptive algorithms with intuitionistic fuzzy logic and forgetting. Educational Technology Research and Development, pp. 1-37. https://doi.org/10.1007/s11423-026-10639-6

[19] Deng, C., Yuan, B. (2026). Research on an intelligent tutoring system based on automatic construction of multimodal knowledge graphs and retrieval-augmented generation. Frontiers in Computer Science, 8: 1777749. https://doi.org/10.3389/fcomp.2026.1777749
[20] Kohl, L., Ansari, F. (2024). A knowledge graph-based learning assistance systems for industrial maintenance. Procedia CIRP, 126: 87-92. https://doi.org/10.1016/j.procir.2024.08.305
[21] Yuanyuan, Y., Ying, Y. (2026). Application and effectiveness analysis of AI intelligent tutoring system combined with knowledge graph in basic medical pathology teaching. BMC Medical Education, 26: 1028. https://doi.org/10.1186/s12909-026-09348-8
[22] Al-Hassan, M., Abu-Salih, B., Alshdaifat, E., Aloqaily, A., Rodan, A. (2024). An improved fusion-based semantic similarity measure for effective collaborative filtering recommendations. International Journal of Computational Intelligence Systems, 17: 45. https://doi.org/10.1007/s44196-024-00429-4
[23] Reig Alamillo, A., Torres Moreno, D., Morales González, E., Toledo Acosta, M., Taroni, A., Hermosillo Valadez, J. (2023). The analysis of synonymy and antonymy in discourse relations: An interpretable modeling approach. Computational Linguistics, 49(2): 429-464. https://doi.org/10.1162/coli_a_00477
[24] Xu, Y., Hu, L., Zhao, J., et al. (2025). A survey on multilingual large language models: Corpora, alignment, and bias. Frontiers of Computer Science, 19(11): 1911362. https://doi.org/10.1007/s11704-024-40579-4
[25] Pakray, P., Gelbukh, A., Bandyopadhyay, S. (2025). Natural language processing applications for low-resource languages. Natural Language Processing, 31(2): 183-197. https://doi.org/10.1017/nlp.2024.33
[26] Rakhimova, D., Turarbek, A., Karyukin, V., Sarsenbayeva, A., Alieyev, R. (2025). Legal AI in low-resource languages: Building and evaluating QA systems for the Kazakh legislation. Computers, 14(9): 354. https://doi.org/10.3390/computers14090354
[27] Ilić, M., Mikić, V., Kopanja, L., Vesin, B. (2023). Intelligent techniques in e-learning: A literature review. Artificial Intelligence Review, 56(12): 14907-14953. https://doi.org/10.1007/s10462-023-10508-1
[28] Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., Du, Z. (2024). Artificial intelligence in education: A systematic literature review. Expert Systems with Applications, 252: 124167. https://doi.org/10.1016/j.eswa.2024.124167
[29] Bi, R. (2025). An adaptive semantic retrieval framework for digital libraries integrating graph neural networks, ontology, and user behavior. Scientific Reports, 15: 40528. https://doi.org/10.1038/s41598-025-24276-1
[30] Ali, R.S., Abouel-Ela, M., Eldakhly, N.M. (2025). An ontology-based adaptive tutoring system for learning business English idioms. Neural Computing and Applications, 37(27): 22725-22753. https://doi.org/10.1007/s00521-025-11506-w
[31] Li, Z., Wang, Z., Wang, W., Hung, K., Xie, H., Wang, F.L. (2025). Retrieval-augmented generation for educational application: A systematic survey. Computers and Education: Artificial Intelligence, 8: 100417. https://doi.org/10.1016/j.caeai.2025.100417
[32] Yilmaz, R., Yurdugül, H., Yilmaz, F.G.K., et al. (2022). Smart MOOC integrated with intelligent tutoring: A system architecture and framework model proposal. Computers and Education: Artificial Intelligence, 3: 100092. https://doi.org/10.1016/j.caeai.2022.100092
[33] Kasneci, E., Sessler, K., Küchemann, S., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103: 102274. https://doi.org/10.1016/j.lindif.2023.102274
[34] Shi, Y., Yu, K., Dong, Y., Chen, F. (2026). Large language models in education: A systematic review of empirical applications, benefits, and challenges. Computers and Education: Artificial Intelligence, 10: 100529. https://doi.org/10.1016/j.caeai.2025.100529
[35] Yan, L., Sha, L., Zhao, L., et al. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. British Journal of Educational Technology, 55(1): 90-112. https://doi.org/10.1111/bjet.13370