International Journal of Innovative Research and Scientific Studies, 8(3) 2025, pages: 2413-2430

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Development of a smart textbook in informatics with interactive learning support and Kazakh language integration based on artificial intelligence technologies

R. Niyazova[1], A. Aktayeva[2*], A. Sharipbay[3], A. Kubigenova[4], B. Razakhova[5]

[1,3,5]*Department of Artificial Intelligence Technologies, L.N.Gumilyov Eurasian National University, Astana 010000, Kazakhstan*.
[2]*Department of Information Systems and Informatics, A. Myrzakhmetov Kokshetau University, Kokshetau 000002, Kazakhstan*.
[4]*Department of Information Systems, S.Seifullin Kazakh Agro Technical Research University, Astana, 010000, Kazakhstan*.

Corresponding author: A. Aktayeva (*Email: aakhtaewa@gmail.com*)

Abstract

This article presents a project for developing interactive models of an educational resource for determining the semantic proximity between the ontology of knowledge extracted from a knowledge base based on a given question and the ontology generated from a student's answer in computer science. A general methodology for developing software has been proposed. This methodology will form the basis for the automated creation of smart textbooks on computer science in the Kazakh language using ontological models and thesauri. These educational resource models are suitable for any type of educational process (Blended Learning Technology (BLT), full-time, part-time). For example, the article describes an online smart textbook that will adapt to the student's individual learning path by providing personalized text, audio, and video materials, asking questions, and evaluating answers with an indication of percentage accuracy. A pedagogical experiment has been conducted to assess the students' performance. The online smart textbook will adapt to the student's learning style by providing personalized text, audio, and video materials. It will also ask questions and evaluate answers with an indication of percentage accuracy. In addition, completed assignments were analyzed to assess students' progress using static digital materials and standard computer testing systems for a dynamic, intelligent learning environment and knowledge assessment. The results show that the proposed methodology has great potential for increasing student engagement in studying the formalization and processing of the grammar of the Kazakh language using production rules and, based on them, developing a grammar processor, creating ontologies, thesauri, and knowledge bases on the content of "Computer Science".

Keywords: Artificial intelligence, Computer linguistics, Computer logic, Knowledge bases, Ontology, Question-answering systems, Smart textbooks, Thesaurus.

DOI: 10.53894/ijirss.v8i3.7020
Funding: This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP19678613).
History: Received: 18 March 2025 / **Revised:** 22 April 2025 / **Accepted:** 24 April 2025 / **Published:** 14 May 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

In recent years, against the backdrop of widespread online distance learning, digital educational resources and tools have actively developed. One of the promising areas in this field is the introduction of smart textbooks—platforms aimed at supporting and distributing electronic educational content, taking into account the structural analogy with traditional printed publications. Smart textbooks are not just digital copies of books but are carefully designed interactive resources that combine text materials, visual and audio elements, and adaptive learning activities to improve knowledge acquisition efficiency.

Artificial intelligence (AI) technologies are significant in developing innovative textbooks. Solutions such as question-answer systems, chatbots, virtual assistants, and competent consultants are examples of AI's ability to support the educational process. The use of AI allows for the creation of an intelligent learning environment that can, in certain aspects, replace the activities of a teacher. According to research by B. Bloom (1984), students who learn under the individual guidance of a teacher using regular assessment and learning correction demonstrate results that are two standard deviations higher than those who learn under the traditional frontal model. This discovery inspired researchers in the field of AI to develop intelligent tutoring systems (ITS) adapted to the individual educational needs of students [1]. The purpose of such systems is not only to provide multimedia educational material but also to create an environment that promotes the development of critical thinking and analytical skills. Intelligent tutoring systems are successfully implemented in various fields of education — from school and university education to the corporate sector and government agencies.

One of the key objectives of Intelligent Tutoring Systems (ITS) is to increase the understanding of student behavior through interactive interaction, which, in turn, contributes to the formation of personalized learning trajectories. Intelligent tutoring systems (ITS) have a significant history of effective research and continue to develop in line with modern pedagogical and technological approaches. One of the earliest and most cited analytical reviews in this area was presented by authors Nwana [2]. In this work, the author defines the main components of ITS and analyzes the evolution of computer-based learning based on a comparison with ITS known at that time.

In Shute and Psotka [3] published a more in-depth study covering the history, evaluation methods, and prospects for further development of intelligent tutoring systems [3]. The authors, Woolf and others, further expand on the topic, paying special attention to ITS's capabilities in interacting with students and demonstrating the potential of systems to understand and adapt to user behavior [4].

Along with general reviews, studies devoted to individual aspects of ITS are of considerable interest. So, Murray [5] compared authoring tools and course design tools in intelligent tutoring systems and what they can accomplish. The thematic reviews by authors Rus et al. [6] reflect an important direction in developing interactive ITS.

The scientific literature pays particular attention to constraint-based learning, the history and development of which Mitrovic discussed in detail in 2012. This approach expands the functionality of the IOS due to the flexible adaptation of training scenarios [7].

Finally, some modern studies Fournier-Viger et al. [8] focus on the features of the ITS's functioning in conditions of uncertainty and weakly structured subject areas and also analyze the effectiveness of systems' behavior in situations of partial or ambiguous formalization of knowledge.

One of the main disadvantages of non-intelligent systems is the passive participation of students. Some studies argue that active involvement and immediate feedback when solving problems and simulations lead to a better perception of information and, accordingly, to better results. Therefore, the primary focus of modern educational platforms is on including interactive simulations, problem-solving exercises, and adaptive teaching methods to meet students' individual needs [9]. In areas of computer science theory that require a deep understanding of abstract mathematical concepts, such as determining the semantic similarity between the ontology of knowledge extracted from a knowledge base based on a given question and the ontology, these educational approaches would be extremely valuable.

The implementation of intelligent tutoring system models of educational resources can provide students with the opportunity to master these concepts in both a visual and practical way. Studies on intelligent tutoring system models indicate that these technologies can help students understand and remember information better, especially when they relate knowledge to similar concepts in a structured way. Intelligent tutoring systems can include points, levels, achievements, and tasks that motivate students to complete them and progress in learning. Integrating problem-solving exercises, automated feedback, and adaptive difficulty levels of educational resources can bridge the gap between theoretical information and practical application.

This paper proposes a general development of a technology for creating smart textbooks capable of interactive teaching, consulting, and assessing a student's knowledge of the Kazakh language with an individual learning trajectory. The methodology provides a systematic, unified approach to developing an innovative textbook that can create a corresponding ontology of student questions and answers, questions and answers from the knowledge base, and questions and answers from the test database, and determine the closeness of the corresponding ontologies by the semantic distance between them.

Unlike traditional static educational materials and passive modeling (*simple animations with limited or no interactivity*), the proposed educational resource model in the field of computer science theory (IOS) integrates modern interactive technologies, which include:

1. A method for generating a knowledge base that describes in the Kazakh language the properties and relationships of concepts included in the ontology and thesaurus of a specific academic discipline. This method formally represents the subject area and serves as a basis for subsequent semantic analysis.
2. . Method for determining an individual student's learning path based on an analysis of the knowledge base. The proposed approach allows for adapting educational content according to the level of training, interests, and results of interaction with the system.
3. Method of semantic comparison of the ontology of the student's question with the ontology of the answer from the knowledge base using fuzzy logic. The technique allows for considering the uncertainty and incompleteness of formulations and assessing the degree of semantic similarity between the posed question and the extracted answer.
4. The fuzzy logic apparatus is used to compare the ontology of the question formed based on the knowledge base with the ontology of the student's answer. This approach provides a more flexible and context-dependent assessment of the quality of the answer, going beyond the formal comparison of key terms.

The methods developed for the project allow for the implementation of an intelligent system capable of semantic analysis of requests and responses, building personalized learning paths, and providing adaptive feedback.

2. Materials and Methods

A methodology for designing and developing a model of an educational resource for the ITS (Intelligent Tutoring Systems) to solve problems in computer science theory. Each of the proposed models of educational resources will address issues as a separate ITS.

This study used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology [10, 11]. The PRISMA methodology encompasses several steps to achieve the desired results, with an information flow similar to that shown in Figure 1.

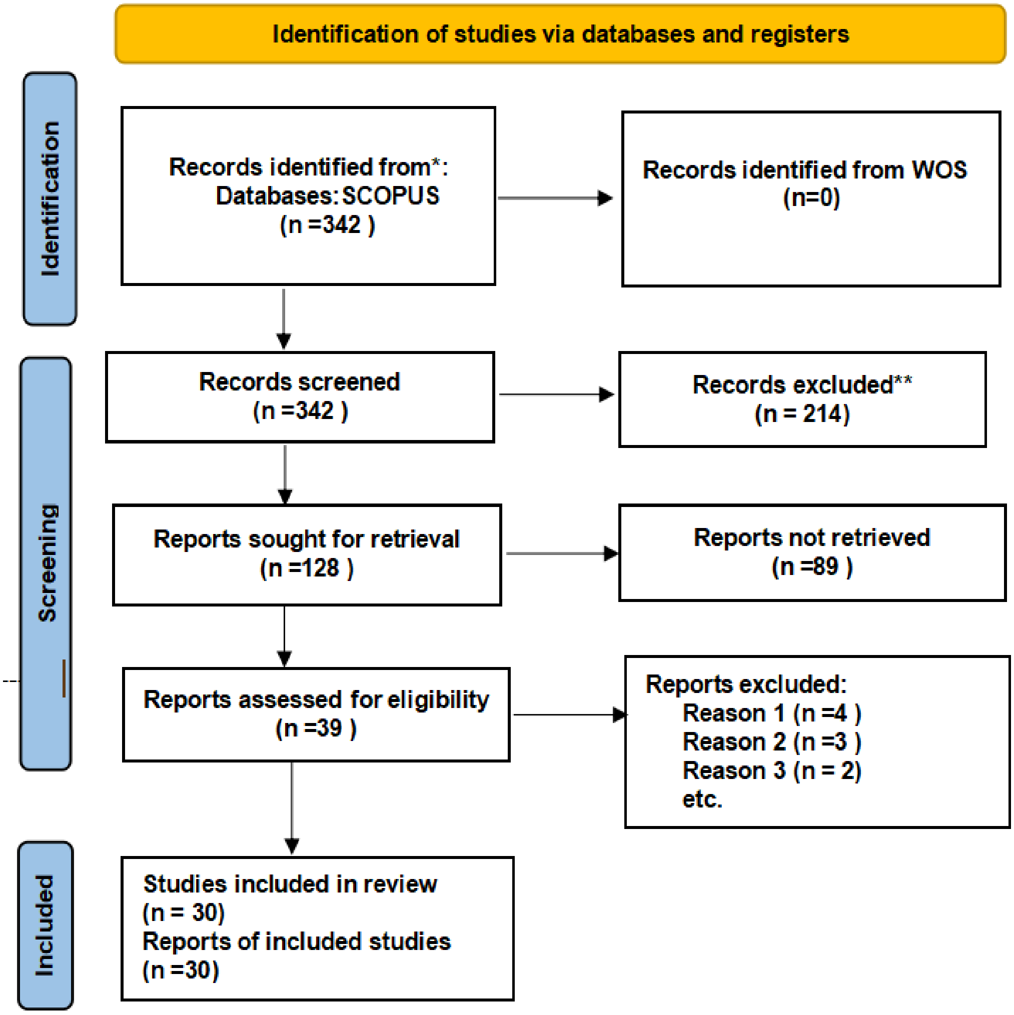

Figure 1.
PRISMA Methodology.

Figure 1 shows the scheme for selecting publications using the PRISMA methodology. Studies were identified from several databases and went through several filtering stages by publication date, document type, language, journal quality, abstract relevance, and duplicate removal. Each stage sequentially narrowed the sample, ensuring that only the highest quality and most relevant publications were included in the final analysis.

The PRISMA framework has been chosen due to its recognized comprehensive nature, wide international application in various scientific fields, and ability to improve the consistency and rigor of systematic reviews [10, 12].

The PRISMA methodology was then applied with step-by-step filtering. The first selection criterion was the year of publication. Furthermore, we used a structured screening and exclusion methodology based on the following criteria:

1. The article does not include an abstract or English text.
2. 1. Inconsistency with the topic of the study (e.g., works not related to ontologies or artificial intelligence).
3. 1. Publications in journals with a low citation index or not included in recognized indices (Scopus Q4, without impact factor, etc.).
4. Duplicate records.
5. Examine articles without original data or methodology.

BD Scopus was used to collect data, which minimized the number of duplicate records. The search query was a logical combination of key terms: "artificial intelligence" AND "computer linguistics" OR "computer logic" AND "ontology" AND "thesaurus," using quotation marks for exact phrase matching. The first stage formed a corpus of 39 records [13-39].

The final sample included top-notch publications on the research topics of creating methods for matching semantic ontologies, using fuzzy logic in educational systems, and building smart learning environments. The initial search yielded 342 results. After filtering studies conducted between 2004 and 2025, 342 documents were identified. Further screening excluded non-empirical studies and articles not in English, resulting in 128 documents. Seven were removed due to retraction, and five were excluded because they were still in press. This step narrowed the sample to 89 papers; the final datasets included 30 studies [13-39] (See Figure 2).

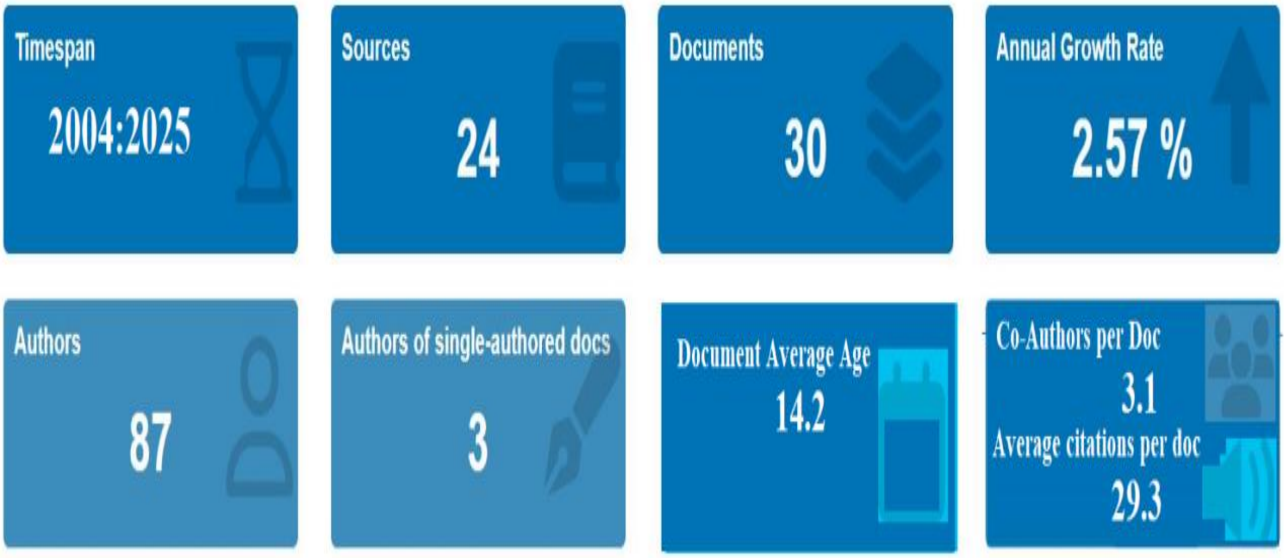

Figure 2.
A statistical graph of the number of studies in different periods.

Figure 3 collects the bibliometric results of the selected articles and can also be analyzed by examining the dots along each axis, representing data values for specific years. The distance from the chart's center indicates the scale of the values, with greater distances indicating more significant values. By comparing changes across the years shown in the diagram, the connecting lines help visualize trends over time.

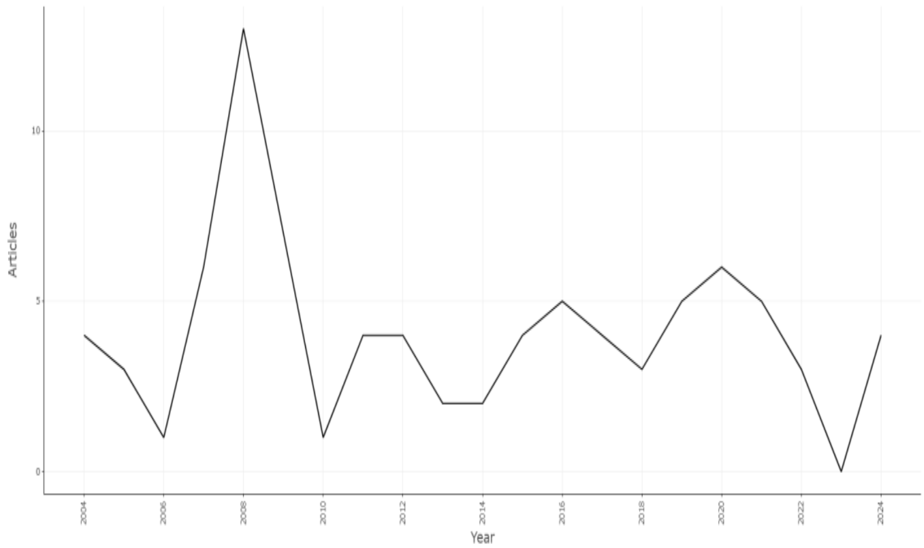

Figure 3.
Bibliometric results of publications from 2004 to 2025.

As shown in Figure 4, the nature of the journals of the identified papers is diverse, covering thesauri and ontology development (Online Information Review), and the most cited is the Journal of Biomedical Informatics. The second most cited [13, 14]. Figure 4 shows that the main field in the selected papers is computer science, followed by mathematics, social sciences, and engineering.

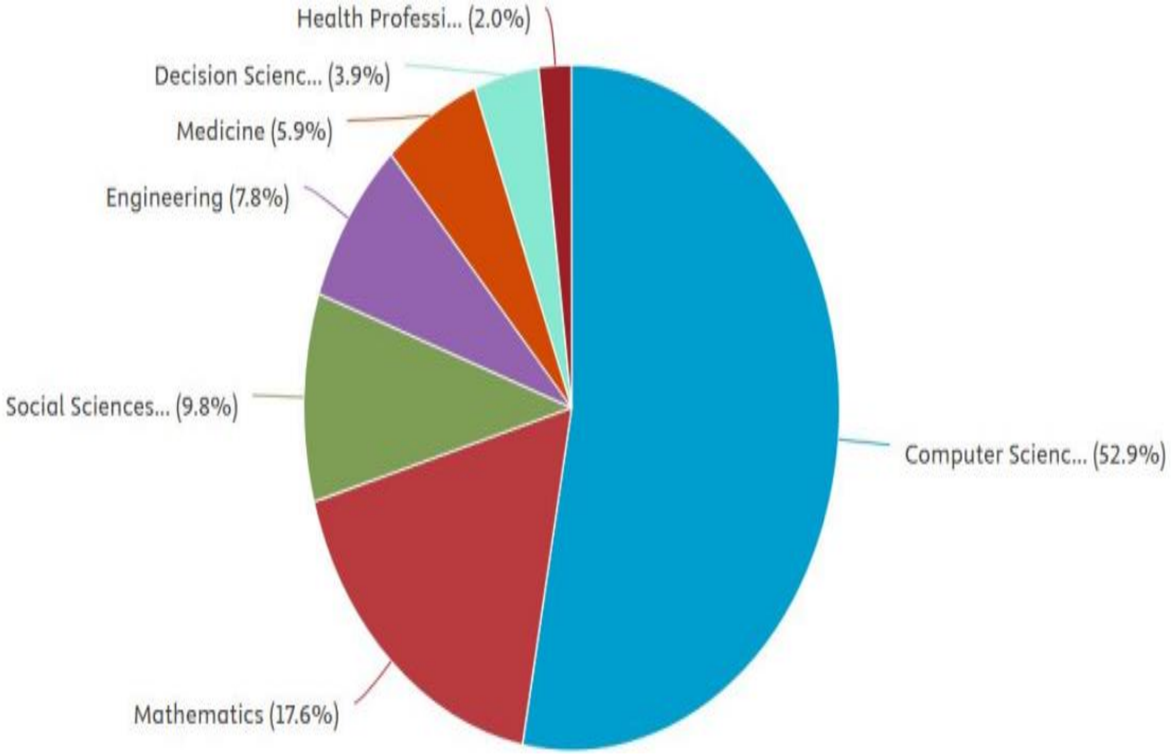

Figure 4.
Bibliometric results: sorted by relevant fields.

Studies flagged as equivocal were further screened using a predetermined decision-making framework to ensure methodological rigor and minimize selection bias. This iterative process strengthened the accuracy and relevance of the final datasets to the review's objectives.

Scientometric Study Using VOSviewer: Key Theme and Relationship Analysis. Scientometric analysis using VOSviewer software (SW) provides valuable information on key research topics, where keywords indicate significant areas of interest. This study employs the co-occurrence method, which treats keywords as units of analysis. The resulting network illustrates the structure of research topics and their relationships.

Co-occurrence analysis determines the relatedness of elements based on the frequency of their co-occurrence in scientific publications. We use a complete count method, assigning equal weight to each occurrence. A minimum occurrence threshold of approximately 5 has been used to extract 372 keywords from the 1048 considered, forming eight clusters with internal relationships and overall connection strength [15-39]. As shown in Figure 5, the keyword network exhibits a high degree of interrelationship.

In SW VOSviewer, clusters are groups of related terms, where each term belongs to only one cluster. Different colors indicate clusters:

Red: “artificial intelligence”, “search engines”, “query languages”, “natural language processing”, “linguistics”, “world wide web”;

Green: “thesauri”, “metadata”, “ontology development”, “information retrieval systems”, “knowledge acquisition”, “text processing”, “algorithms”;

Blue: “question answering”, “computational linguistics”, “semantic web”, “natural language processing”, “data handling”;

Yellow: “ontology”, “data description”, “knowledge management”, “semantics”, “domain ontologies”.

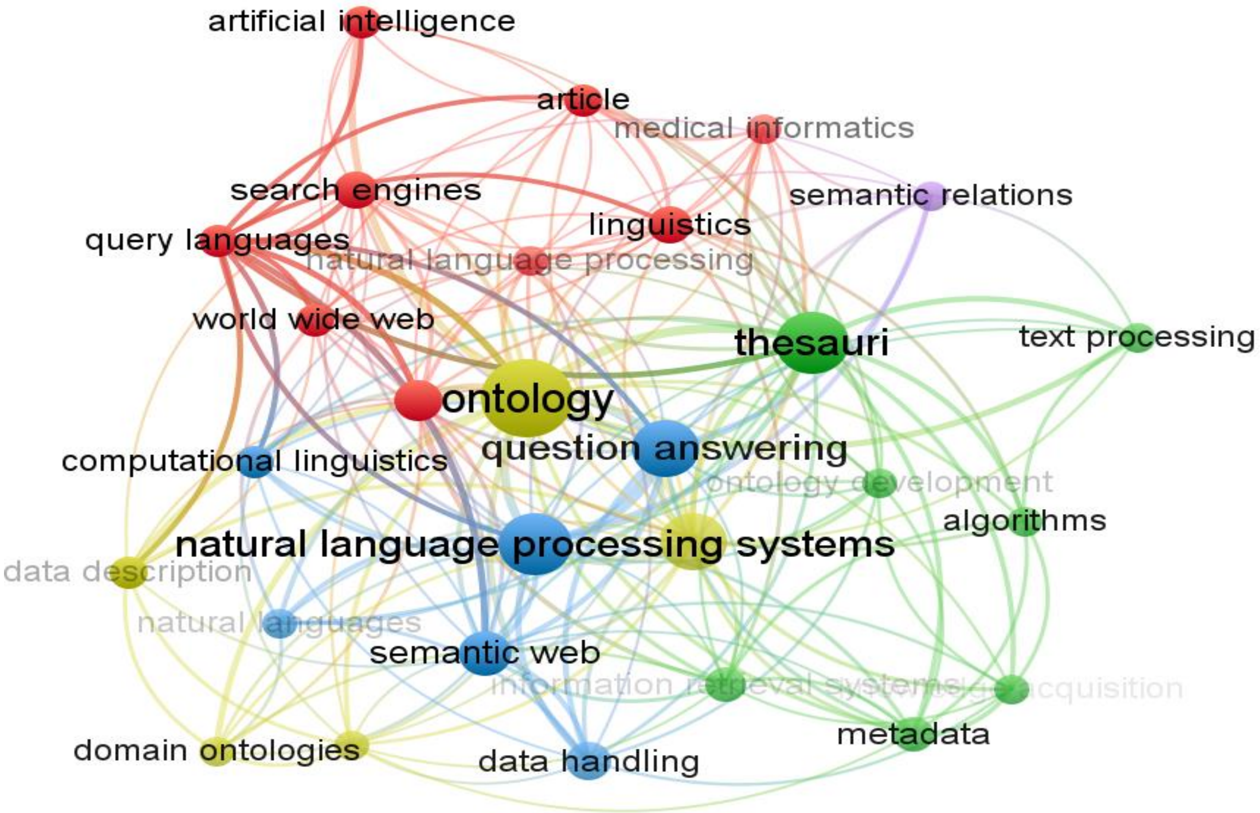

Figure 5.
Co-occurrence of keywords in VOSviewer.
Note: The circled node has been selected in Figure 9.

Relationships between terms are represented by lines whose thickness reflects the strength of the relationship: the thicker the line, the higher the frequency of co-occurrence of the corresponding terms. For example, terms such as “ontology,” “question answering,” “natural language processing systems,” and “thesauri” demonstrate strong relationships.

Figure 6 shows that when a particular node (term) is selected, its related elements are automatically highlighted, indicating their importance and position in the topic field's structure.

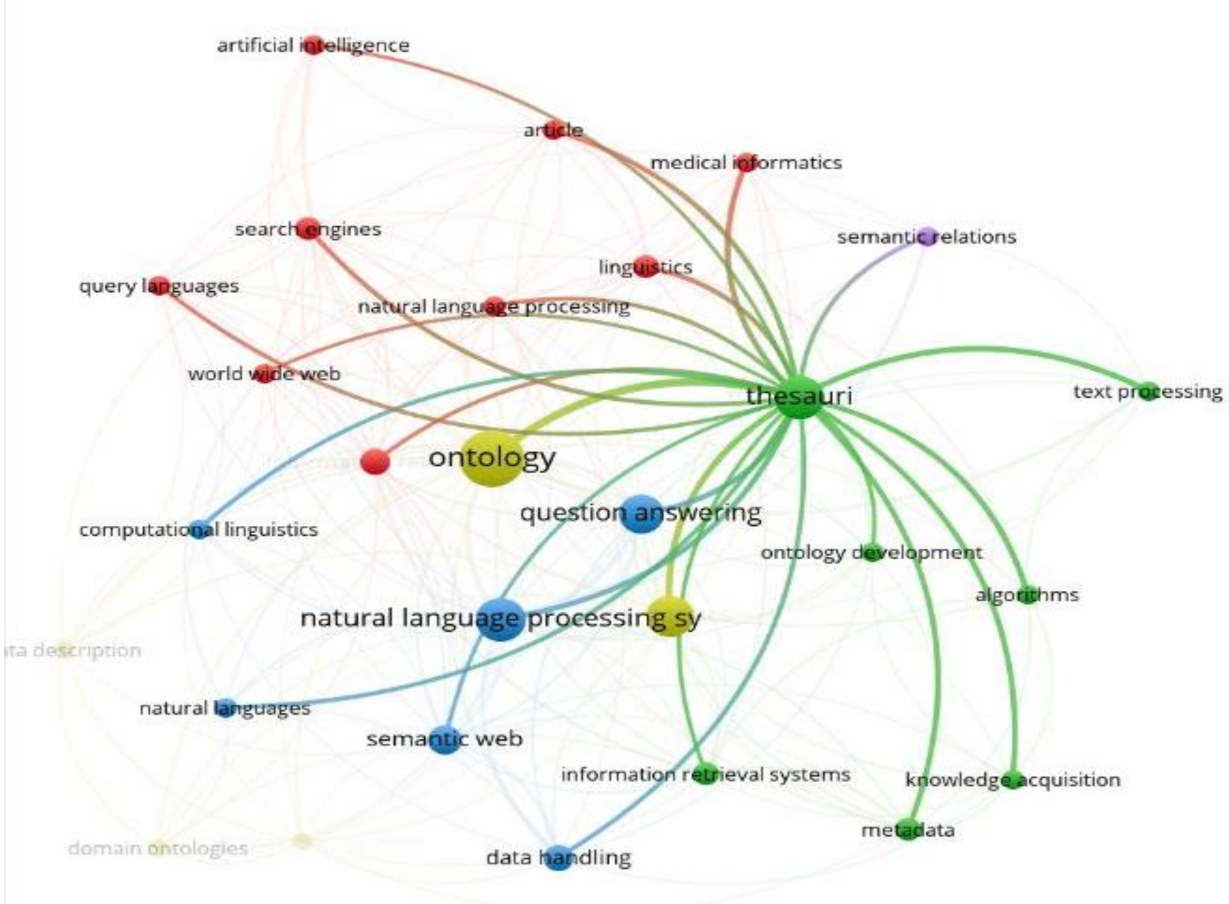

Figure 6.
Selecting a node in VOSviewer.

Key observations include the early prominence of foundational topics such as “artificial intelligence,” “search engines,” “query languages,” “World Wide Web,” “medical informatics,” “natural language processing,” “knowledge acquisition,” and “algorithms,” which are central to the blue regions.

The analysis reveals key trends, including the early dominance of fundamental themes such as artificial intelligence, search engines, query languages, the World Wide Web, medical informatics, natural language processing, knowledge acquisition, and algorithms. These themes are central to the blue areas of the web and lay the foundation for subsequent developments in the field.

The evolution of Research Areas and Technology Integration. As time passes, as reflected by the network's shift to green and yellow shades, new research areas emerge, including question answering, ontology, text processing, computational linguistics, semantic relations, data handling, knowledge management, and natural language processing. This shift indicates a growing interest in integrating advanced technologies, artificial intelligence, and sustainable solutions.

The yellow nodes show ideas such as knowledge management, semantic relations, natural language processing, and data handling, pointing out trends in the growth and use of intelligent systems. Researchers are currently concentrating on developing adaptive educational technologies, which encompass the creation of smart textbooks. These learning resources can analyze student questions, create customized learning paths, and automatically evaluate answers based on a test database. Figure 7 shows a network visualization combined with a gradient bar representing a timeline from 2004 to 2024, illustrating the temporal evolution of research focus and the emergence of new areas in the field [13-39]. Nodes of colors closer to blue represent earlier research trends (closer to 2008), while green and yellow nodes reflect more recent developments (up to 2025).

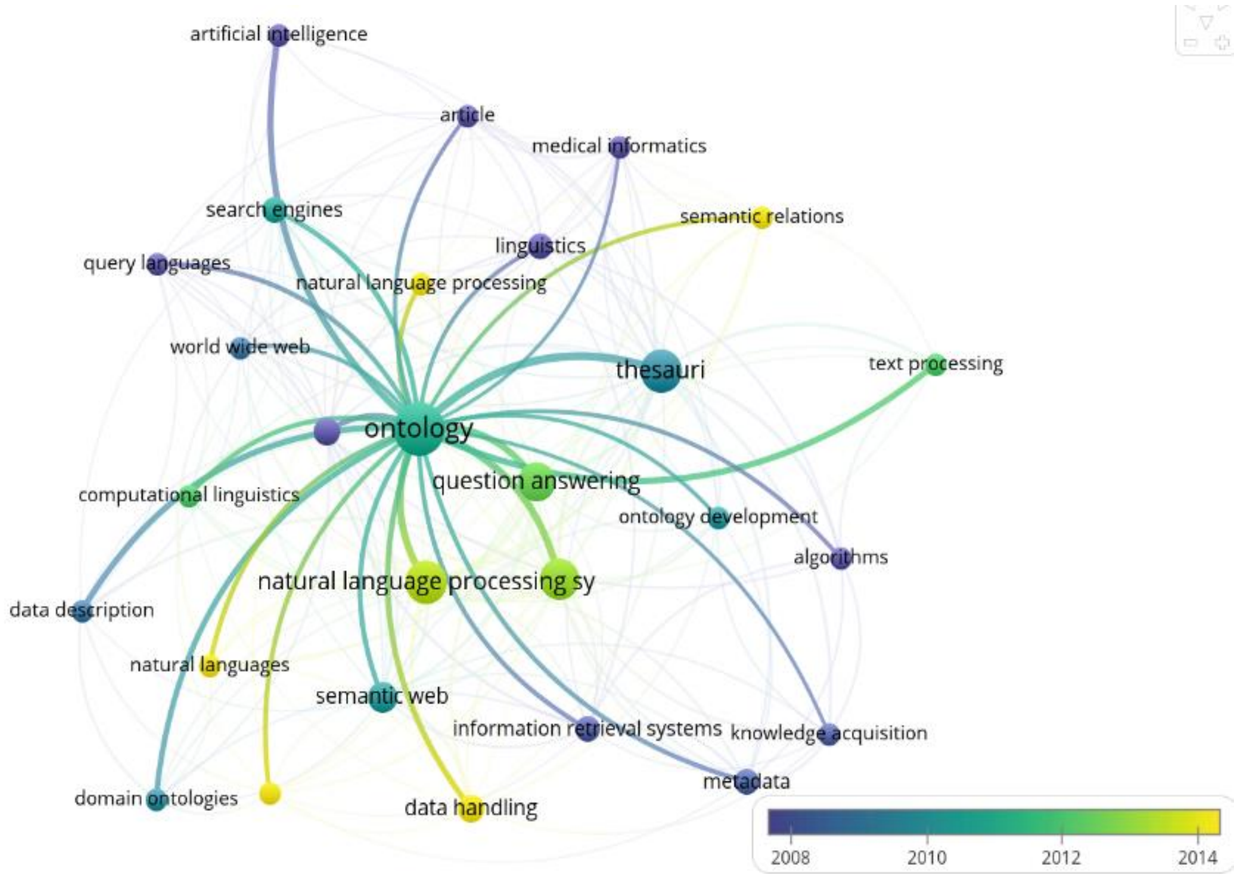

Figure 7.
Keyword Sharing in VOSviewer.

Specifically, we apply the technology to "Computer Science" and develop a specialized Internet portal. This portal includes a user interface that can conduct written and oral dialogue in the Kazakh language, controlled by artificial intelligence. Thus, the visualization reflects a transition from traditional approaches to intelligent educational systems and innovative solutions in the field of knowledge. In this regard, it is advisable to formulate a general methodology for their construction. The methodology for developing a model of the educational resources of the ITS has been developed. The individual stages of the method are as follows (see Figure 8).

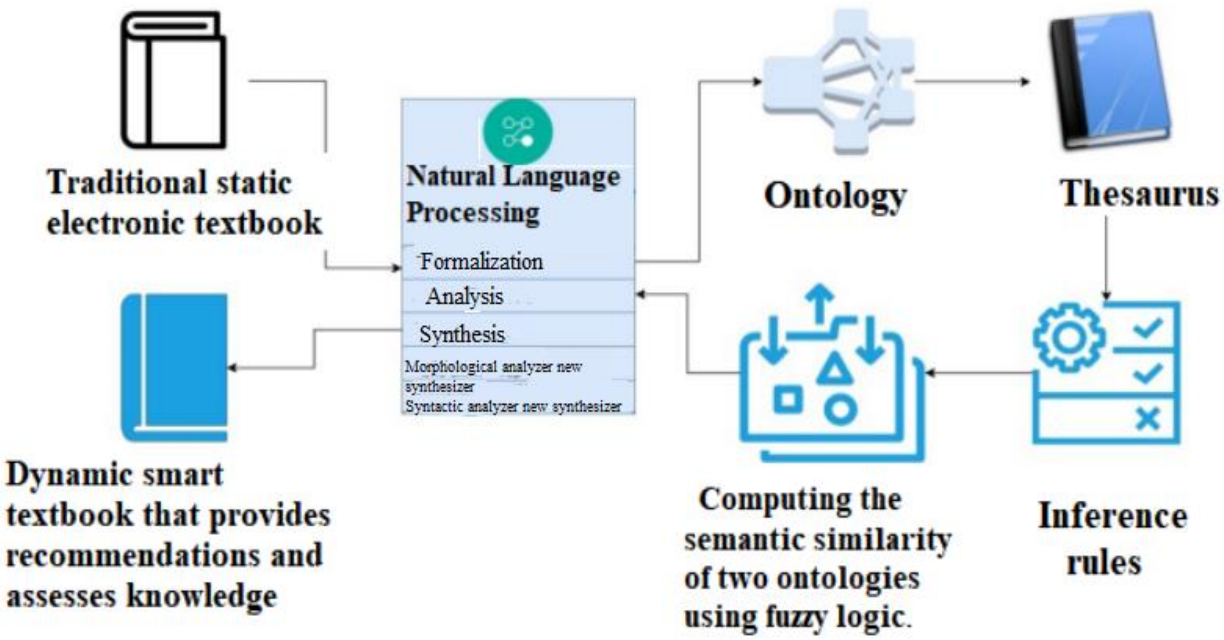

Figure 8.
Methodology for developing a model of the educational resources of the ITS.

The model of the educational resource of the ITS uses a structured and comprehensive approach aimed at ensuring high-quality implementation of the ITS throughout its life cycle, which is a key point in implementing an innovative textbook in the educational environment.

3. Results

Development of a model of the educational resource of the ITS. Within the framework of this study, comprehensive approaches were used to build a methodological model of an intelligent, innovative textbook, including methods of ontological modeling, semantic analysis, and linguistic formalization. The primary methods used are presented below:

1. 1. *Ontological modeling and knowledge base design.* Methods of ontological modeling and knowledge base construction formalize the structure and content of the academic discipline. This ensures a formalized, machine-readable representation of the subject area necessary for intelligent educational information processing.

2. 1. *Methods for assessing the semantic closeness of ontologies*. Algorithms for comparing two ontologies based on fuzzy logic approaches have been developed:

a) a) When comparing the ontology of a question formulated by a student with the ontology of an answer retrieved from the knowledge base of a smart textbook, the correctness of the system's automatic answer is assessed.

b) a) The quality of a student's answer is assessed by comparing the ontology of a question from the knowledge base with the ontology of an answer given by a student. Semantic weighting and fuzzy interpretation of semantic correspondences are used in both cases.

3. *Software methods for constructing an individual educational trajectory*. To implement adaptive learning, a module forms an individual trajectory based on the knowledge base and the results of the student's previous interactions with the system. This module also provides interactive support for learning and automatic knowledge assessment.

4. 1. *Methods of artificial intelligence and digital content generation.* Artificial intelligence and software generation methods automatically create smart textbooks based on the ontologies and thesauri of academic disciplines. This enables the scalable and context-dependent formation of educational content.

5. *Methods for formalizing morphological and syntactic rules of the Kazakh language.* To develop a user interface in the Kazakh language, specialized analyzers (parsers) and synthesizers based on formalized rules of morphology and syntax are implemented.

These components ensure the correct generation and interpretation of texts in natural language within the user's interaction with the innovative textbook. The presented methodological complex allows for implementing an intelligent educational system capable of adapting to the student's characteristics, interpreting and evaluating the meaning of educational requests and responses, and providing personalized recommendations in the learning process.

Addressing Research Questions. This study used complex approaches to build a methodological model of an intelligent smart textbook, including ontological modeling, semantic analysis, and linguistic formalization. The primary methods are presented below:

1. 1. *Ontological modeling and knowledge base design.* Methods of ontological modeling and knowledge base construction have been used to formalize the structure and content of the academic discipline. This ensures a formalized, machine-readable representation of the subject area necessary for intelligent educational information processing.

As part of this study, an ontology of the subject "Computer Science" was developed for the school course from grades 5 to 11. The figure shows a fragmented ontology corresponding to the content of the computer science course for grade 5, created within the software “Protégé” environment using the OntoGraf graphical extension, which allows visualizing all conceptual connections between entities (See Figure 9).

The ontology was developed based on an analysis of the content of a typical educational program. It aims to formalize the subject area for subsequent use in intelligent learning systems and smart textbooks, as well as for automated processing and evaluation of educational information.

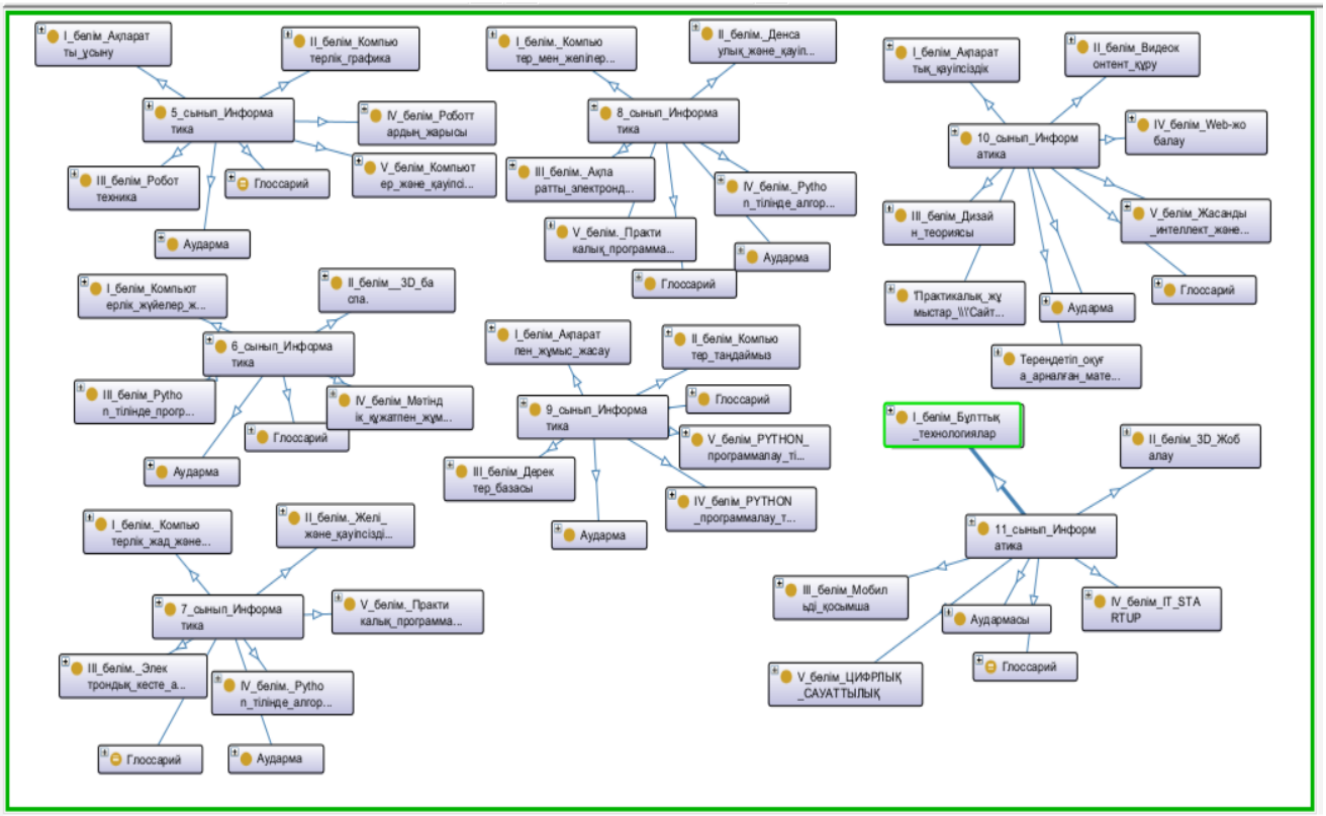

Figure 9.
An ontology of the subject "Computer Science”.

The structure of the ontology covers five key thematic blocks:

A. Representation of information: includes basic concepts related to data types, their presentation and transformation, coding methods, encryption, and information protection.

B. Computer graphics: covers both theoretical and practical aspects of working with raster and vector images, including the use of graphic editors and the execution of relevant tasks.

C. Robotics: contains concepts related to the design elements and operating principles of robots, as well as the basics of their programming.

D. Robot competitions: describe the types and formats of competitions, stages of development and configuration of robots, as well as evaluation criteria and platforms used in educational practice.

E. Computer and safety: includes issues related to the safe use of computing equipment, digital ethics, and basic aspects of information security.

Additionally, a glossary of terms and their multilingual translations was included in the ontological model, ensuring semantic interoperability and expanding the possibilities of application in a multilingual educational environment.

The developed ontology forms a holistic conceptual model of the subject area of school informatics and can be used to implement adaptive learning, build intelligent decision support systems, as well as in tasks of semantic search and analysis of educational data. One key methodological foundation for developing an intelligent textbook is the ontological modeling of the academic discipline. This approach ensures a formalized, machine-readable representation of the subject area necessary for the intelligent processing of educational information and the construction of a knowledge base.

The formation of an ontology involves structuring terms and concepts in a hierarchy, defining semantic relationships between concepts, and ensuring the logical consistency of the model. This knowledge representation format allows for the effective use of methods of semantic inference, automatic search, comparison, and adaptive generation of educational content.

The process of constructing an ontology and knowledge base includes the following stages, presented in the form of pseudocode (See Algorithm 1).

Algorithm 1: Building an ontology of an academic discipline.

```
PROCEDURE BuildOntology (CURRICULUM)
// Step 1. Initialise the ontology structure
ONTOLOGY ← Create_New_Ontology ()
THESAURUS ← Import_Terms (CURRICULUM)

// Step 2. Define key concepts
FOR each TERM in THESAURUS:
CONCEPT ← Create_Concept (TERM)
ADD_TO_ONTOLOGY(CONCEPT)
```

```
// Step 3. Establish hierarchies and relationships
For each PAIR (K1, K2) in CONCEPTS:
IF There_Is_Relationship (K1, K2):
RELATIONSHIP ← Define_Relationship_Type (K1, K2)
ADD_RELATIONSHIP (K1, K2, CONNECTION)

// Step 4. Verification of structure
IF Logical_Consistency_Check (ONTOLOGY) == SUCCESSFUL:
SAVE(ONTOLOGY)
ELSE:
OUTPUT_ERROR ("Violation of logic or cyclicity in ontology")
END OF PROCEDURE
```

Methods for assessing the semantic closeness of ontologies. Algorithms for comparing two ontologies based on fuzzy logic approaches are developed:

a) When comparing the ontology of a question formulated by a student with the ontology of an answer extracted from the knowledge base of an innovative textbook, the correctness of the system's automatic answer is assessed.

b) The quality of a student's answer is assessed by comparing the ontology of a question from the knowledge base with the ontology of an answer given by a student. Semantic weighting and fuzzy interpretation of semantic correspondences have been used in both cases (See Algorithm 2).

Algorithm 2: Semantic similarity evaluation.

```
// Pseudocode of the semantic similarity evaluation algorithm:
PROCEDURE EvaluateSemanticSimilarity(ONTOLOGY_1, ONTOLOGY_2)
CONCEPTS_1 ← Extract_Concepts(ONTOLOGY_1)
CONCEPTS_2 ← Extract_Concepts(ONTOLOGY_2)

OVERALL_Similarity ← 0
COUNT ← 0

FOR each TERM_1 in CONCEPTS_1:
MAX_Similarity ← 0
FOR each TERM_2 in CONCEPTS_2:
SIMILARITY ← CalculateSemanticScore(TERM_1, TERM_2)
IF SIMILARITY > MAX_Similarity:
MAX_Similarity ← SIMILARITY
END FOR
OVERALL_SIMILARITY ← OVERALL_SIMILARITY + MAX_SIMILARITY
COUNT ← COUNT + 1
END FOR

AVERAGE_SIMILARITY ← OVERALL_SIMILARITY / COUNT
RETURN FuzzyInterpretation(AVERAGE_SIMILARITY)
END PROCEDURE

FUNCTION CalculateSemanticScore(TERM_1, TERM_2)
PATH_1 ← Ontology_Path(TERM_1)
PATH_2 ← Ontology_Path(TERM_2)
DISTANCE ← Compare_Lengths(PATH_1, PATH_2)
WEIGHT ← Evaluate_Contextual_Relevance(TERM_1, TERM_2)
RETURN FuzzyScore(DISTANCE, WEIGHT)
END OF FUNCTION
```

An example of the application of the method of assessing the semantic closeness of ontologies. Let us consider a specific example of evaluating the semantic closeness between two ontologies in the context of the discipline "Computer Science" in the Kazakh language.

Suppose a student asks the question:

“Aqparatty kodtau degenimiz ne?” - "What is the meaning of the word code?"

The system retrieves the following related answer from the knowledge base:

“Kodtau — aqparatty arnayy alfavitpen türlendiru protsesi. “ - "The word code is the alphabetic sequence of processes."

To analyze the semantic correspondence between the question and the answer, the following ontological representations are built:

I. Ontology of the question:

1. Concepts: Akparat (information), kodtau (coding).
2. Relationship: kodtau ⊂ akparattyk uderister (information processes).

II. Ontology of the answer:

3. Concepts: kodtau, turlendiru (transformation), alphabet.
4. Relationship: kodtau → turlendiru → alphabet.

III. Steps of assessing semantic similarity.

IV. Extraction of terms:
Tq = {akparat, kodtau},
Ta = {kodtau, turlendiru, alphabet}

Table 1.
Calculation of pair similarities.

Term (question)	Term (answer)	Semantic similarity
Akparat - information	Turlendiru - conversion	0.4
Akparat - information	Kodtau -encoding	0.7
Akparat - information	Kodtau - encoding	1.0

1. Maximum correspondences (by fuzzy logic): information ↔ encoding = 0.7; encoding ↔ encoding = 1.0

2. Aggregate assessment: $\boldsymbol{Average\ value\ =\ \frac{0.7+1.0}{2} = 0.85}$

3. Some interpretation:

According to the given scale:
- 0.0 – 0.4 — low compliance
- 0.4 – 0.7 — average correspondence
- 0.7 – 1.0 — high correspondence

The semantic correspondence between the ontology of the question and the answer is high (0.85). Therefore, the system interprets the automatic answer as correct and relevant.

2. *Program methods of building an individual educational trajectory.* To implement adaptive learning, a module has been developed that forms an individual trajectory based on the knowledge base and the results of the learner's previous interaction with the system. This module also provides interactive learning support and automatic knowledge assessment (See Algorithm 3).

Algorithm 3: Procedure of adaptive training.

```
PROCEDURE SmartTextbookSystem()
INITIALIZE Modules (NLP, Logic, Ontology, Q&A, Adaptive Learning)

WHILE learning process is active
INPUT <- Get input from student

IF INPUT == Question
TEXT_ANALYSIS <- NLP_Module(INPUT)
MEANING_CONNECTIONS <- LogicOntologyModule(TEXT_ANALYSIS)
RESPONSE <- QASystem(MEANING_CONNECTIONS)

ELSE IF INPUT == Test
STUDENT_ANSWERS <- Get answers
EVALUATE <- EvaluateStudent(STUDENT_ANSWERS, CORRECT_ANSWERS)
CORRECT_LEARNING <- AdaptiveLearning(RATING)

ELSE
SmartTextbook <- Offer explanation
OUTPUT ANSWER
END BYE
END PROCEDURE

FUNCTION NLP_Module(TEXT)
TEXT <- Strip punctuation, normalise
TOKENS <- Break text into words
MEANING <- Analyse syntax and semantics
RETURN MEANING
END FUNCTION

FUNCTION LogicOntologyModule(INPUT_DATA)
```

```
IF INPUT_DATA matches known ontology
RESULT <- Find relationships and concepts
ELSE
RESULT <- Query thesaurus to clarify terms
RETURN RESULT
END FUNCTION

FUNCTION QASystem(QUESTION)
MEANING <- NLP_Module(QUESTION)
INFO <- LogicOntologyModule(MEANING)
ANSWER <- Find in knowledge base
IF ANSWER == NULL
ANSWER <- Generate answer via GPT
RETURN ANSWER
END OF FUNCTION

FUNCTION EvaluateStudent(ANSWERS, CORRECT_ANSWERS)
CORRECT <- 0
BYPASS EACH ANSWER INTO ANSWERS
IF ANSWER == CORRECT
CORRECT <- CORRECT + 1
SCORE <- (CORRECT / TOTAL_NUMBER) * 100
RETURN SCORE
END OF FUNCTION

FUNCTION AdaptiveLearning(RESULTS)
IF RESULTS < 50%
CONCLUSION: "We recommend repeating the theory."
ELSE IF RESULTS < 80%
CONCLUSION: "Try to complete additional exercises."
ELSE
CONCLUSION: "You can proceed to the next level."
END OF FUNCTION
```

Methods of artificial intelligence and generation of digital content. Artificial intelligence and software generation methods are applied to create smart textbooks automatically based on the ontologies and thesauri of educational disciplines. This allows for the scalable and context-dependent formation of educational content.

One key direction in developing intellectual educational systems is automating the generation of educational materials. As part of the proposed architecture of the smart textbook, methods of artificial intelligence and software content generation based on ontologies and thesauri of educational disciplines are used. This approach provides scalable, flexible, and context-dependent formation of digital educational content.

Educational materials are developed based on a structured subject area (ontology) and a terminological base (thesaurus), combined with the capabilities of generative language models. Below is an algorithm that illustrates the process of the automated generation of learning blocks (see Algorithm 4).

Algorithm 4: Procedure for generating educational content.

```
PROCEDURE GenerateSmartContent()
// Initialize components
Load Domain_Ontology
Load Subject_Thesaurus
Load Language_Model (e.g. GPT)

FOR each Section in Subject_Ontology:
Key_Terms ← Extract_Concepts(Section, Thesaurus)
Semantic_Relationships ← Build_Context(Key_Terms, Ontology)
// Generate learning components
Explanatory_Text ← GPT.Generate_Explanation(Semantic_Relationships)
Test_Questions ← GPT.Create_Questions(Key_Terms)
Visual_Materials ← Visualization_Generator(Key_Terms)

// Saving training materials
Save(Explanatory_Text, Test_Questions, Visual_Materials)
END FOR
```

```
END OF PROCEDURE
```

3. 1. *The method of formalization of morphological and syntactic rules of the Kazakh language.* To develop a user interface in the Kazakh language, specialized analyzers (parsers) and synthesizers based on formalized rules of morphology and syntax are implemented. These components ensure the correct generation and interpretation of texts in natural language within user interaction with the smart textbook (See Algorithm 5.1 and 5.2).

Algorithm 5.1: Morphological-syntactic analysis.

```
PROCEDURE KazakhLanguageProcessor(INPUT_TEXT)
CLEANED_TEXT ← Clear_punctuation(INPUT_TEXT)
TOKENS ← Split_into_words(CLEANED_TEXT)

FOR each WORD in TOKENS:
STEM ← Find_stem(WORD)
AFFIXES ← Highlight_affixes(WORD)
PART_OF_SPEECH ← Determine_part_of_speech(STEM, AFFIXES)
ADD to MORPHSULT: (STEM, AFFIXES, PART_OF_SPEECH)
END FOR

SYNTH_STRUCTURE ← Build_syntax_tree(MORPH_RESULT)
RETURN SYNT_STRUCTURE
END OF PROCEDURE
```

Algorithm 5.2: Generating a sentence in Kazakh.

```
PROCEDURE KazakhSentenceGenerator(SEMANTIC_DATA)
SUBJECT ← Generate_subject(SEMANTIC_DATA)
PREDICT ← Generate_verb(SEMANTIC_DATA)
ADDITIONS ← Add_context_modifiers()

SENTENCE ← Assemble_in_order (SUBJECT, ADDITIONS, PREDICT)
SENTENCE ← Apply_morphological_endings(SENTENCE)
RETURN OFFER
END OF PROCEDURE
```

To ensure the full functionality of the intelligent smart textbook in the Kazakh language, it is necessary to implement language modules capable of processing natural user statements, interpreting their requests, and generating correct responses. For this purpose, specialized analyzers and synthesizers based on the Kazakh language's formalized morphological and syntactic rules are integrated into the system architecture.

The morphological analysis allows for identifying word stems, affixes, and grammatical categories, which are critically important for languages with an agglutinative structure, such as Kazakh. Syntactic analysis is necessary to determine the structure of a sentence, including the subject, predicate, and adverbial components. Text generation in the Kazakh language also requires compliance with grammatical and morphological norms corresponding to the semantic context. Below is a generalized pseudocode demonstrating the principles of operation of linguistic modules (See Algorithm 6).

Algorithm 6: The principles of operation of linguistic modules

```
PROCEDURE KazakhLanguageProcessor(INPUT_TEXT)
// Stage 1: Preprocessing
CLEANED_TEXT ← Remove_punctuation(INPUT_TEXT)
WORDS ← Tokenize(CLEANED_TEXT)

// Stage 2: Morphological analysis
FOR each WORD in WORDS:
STEM ← Find_stem(WORD)
AFFIXES ← Determine_affixes(WORD)
PART_OF_SPEECH ← Classify(STEM, AFFIXES)
ADD to MORPHS: (STEM, AFFIXES, PART_OF_SPEECH)
END FOR

// Stage 3: Syntactic analysis
SYNTAX ← Build_sentence_structure(MORPH_PARS)
RETURN SYNTAX
END OF PROCEDURE

PROCEDURE KazakhTextGenerator(SEMANTIC_CONTEXT)
```

```
SUBJECT ← Generate_subject(by_context)
VERB ← Select_action(by_meaning)
ADDITIONS ← Add_advancements(by_context)

SENTENCE ← Assemble(SUBJECT, ADDITIONS, VERB)
SENTENCE ← Apply_morphology(SENTENCE)
RETURN SENTENCE
END OF PROCEDURE
```

The presented methodological complex allows for the implementation of an intelligent educational system capable of adapting to the individual characteristics of the student, interpreting and evaluating the meaning of educational requests and responses, as well as providing personalized recommendations in the learning process:

1. KazakhLanguageProcessor — performs linguistic analysis: morphemic analysis (base + affixes), classification by parts of speech, and syntactic linking of words.
2. KazakhSentenceGenerator — synthesizes sentences considering the structure of the Kazakh language (subject-verb-object), applying the rules of agreement and agglutination.

4. Discussion

Modern research in intelligent teaching systems (ITS) emphasizes their potential for adapting to students' individual characteristics. ITS that simulates interaction with a human tutor plays a special role, especially when learning is difficult. Visualization and modeling (for example, geometry and programming) help improve the quality of education.

The transition of ITS from static to flexible and context-sensitive systems reflects modern trends in digital learning. Future research should improve dialogue mechanisms, expand subject areas, and integrate with AI technologies.

The project includes building databases and knowledge, modeling, programming, and testing. The activities are organized into three groups:

1. Study of innovative textbooks and requirements for them - analysis of international and domestic standards, as well as the creation of text and audio corpora on the content of the subject "Computer Science."
2. Formalization of grammar and linguistic base—development of a grammar processor for the Kazakh language based on production rules, creation of ontologies, thesauri, and knowledge bases on the subject of "Computer Science."
3. The development of an innovative textbook generation technology—creating a textbook capable of answering questions, evaluating student answers, and developing an Internet portal with a dialogue interface in the Kazakh language.

Thus, the project combines theoretical research and practical developments to create intelligent educational solutions that can be implemented in the educational process.

5. Conclusions

The developed models and methods of formalization, presentation, and processing of knowledge contribute to developing both theoretical and applied aspects of artificial intelligence and information technology. The fundamental difference between the project idea and existing analogues is the development of a technology for generating innovative textbooks based on artificial intelligence methods and tools capable of conducting written and oral online dialogue in Kazakh.

The project's results include the technology for generating a smart textbook, a smart textbook created in the Kazakh language, and an Internet portal that provides for this textbook and related services; an ontological model of the subject area, thesauri, and knowledge bases; and methods for determining the semantic proximity of ontologies. The innovative textbook can conduct written and oral dialogue in the Kazakh language online to determine an individual learning path, conduct online consultations, and assess knowledge.

Implementing the project results in Kazakhstan's educational system will improve the population's education level, accelerate the learning process, and raise the competitiveness of domestic education to a high international level. The innovative textbook can be used to develop intelligent systems in education, medicine, law, e-government, etc. As a result, the productivity of trained and retrained specialists will increase, which, in turn, will decrease the cost of products and technologies. The project's implementation will improve the quality and image of Kazakhstan's education. All this will achieve high socio-economic results. The following main contributions can be highlighted:

1) Methods of ontological modeling and design of knowledge bases used to formalize the structure and content of the subject of secondary education and build a knowledge base;
2) Methods for assessing the semantic closeness between two ontologies related to the subject: if the semantic closeness is measured between the ontology of the student's question and the ontology of the answer obtained from the smart textbook knowledge base, then the smart textbook's answer is assessed; if the closeness is measured between the ontology of the innovative textbook's question and the ontology of the student's answer, then the student's answer is assessed;
3) Programming methods used to build an individual learning path for a student based on the knowledge base for the subject of Computer Science, implementing interactive learning and assessing knowledge;
4) Methods of artificial intelligence and programming necessary for developing the technology for the automatic generation of an innovative textbook;

5) Methods for formalizing the morphological and syntactic rules of the Kazakh language, as well as creating corresponding analyzers (parsers) and generators (synthesizers) used in developing the user interface of an innovative textbook in the Kazakh language.

References

[1] B. S. Bloom, "The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring," *Educational Researcher,* vol. 13, no. 6, pp. 4-16, 1984. http://dx.doi.org/10.3102/0013189X013006004

[2] H. S. Nwana, "Intelligent tutoring systems: An overview," *Artificial Intelligence Review,* vol. 4, no. 4, pp. 251-277, 1990. https://doi.org/10.1007/BF00168958

[3] V. J. Shute and J. Psotka, *19. Intelligent tutoring systems: Past, present, and future* (Handbook of research for educational communications and technology). Arlington, VA: Macmillan, 1994.

[4] B. P. Woolf, J. Beck, C. Eliot, and M. Stern, *Growth and maturity of intelligent tutoring systems: A status report. In Smart Machines in Education*. Cambridge, MA: MIT Press, 2001.

[5] T. Murray, *An overview of intelligent tutoring system authoring tools: Updated analysis of the state of the art. In Authoring tools for advanced technology learning environments*. Springer. https://doi.org/10.1007/978-94-017-0819-7_17, 2003.

[6] V. Rus, S. D'Mello, X. Hu, and A. Graesser, "Recent advances in conversational intelligent tutoring systems," *AI Magazine,* vol. 34, no. 3, pp. 42-54, 2013. https://doi.org/10.1609/aimag.v34i3.2485

[7] A. Mitrovic, "Fifteen years of constraint-based tutors: What we have achieved and where we are going," *User Modeling and User-Adapted Interaction,* vol. 22, no. 1-2, pp. 39-72, 2012. https://doi.org/10.1007/s11257-011-9105-9

[8] P. Fournier-Viger, R. Nkambou, and E. M. Nguifo, "Building intelligent tutoring systems for ill-defined domains," *Advances in intelligent tutoring systems,* pp. 81-101, 2010. https://doi.org/10.1007/978-3-642-14363-2_5

[9] K. VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems," *Educational Psychologist,* vol. 46, no. 4, pp. 197-221, 2011. https://doi.org/10.1080/00461520.2011.611369

[10] R. L. Costa, Á. L. da Dias, R. Gonçalves, L. Pereira, and S. Abreu, "Artificial intelligence in project management: Systematic," *International Journal of Technology and Intelligent Planning,* vol. 13, pp. 143–163, 2022. https://doi.org/10.1504/IJTIP.2022.10050400

[11] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and P. Group, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *International Journal of Surgery,* vol. 8, no. 5, pp. 336-341, 2010. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

[12] A. Liberati *et al.*, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration," *PLoS Med,* vol. 339, 2009. https://doi.org/10.7326/0003-4819-151-4-200908180-0013

[13] N. Sioutos, S. de Coronado, M. W. Haber, F. W. Hartel, W.-L. Shaiu, and L. W. Wright, "NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information," *Journal of Biomedical Informatics,* vol. 40, no. 1, pp. 30-43, 2007. https://doi.org/10.1016/j.jbi.2006.02.013

[14] F. W. Hartel, S. de Coronado, R. Dionne, G. Fragoso, and J. Golbeck, "Modeling a description logic vocabulary for cancer research," *Journal of Biomedical Informatics,* vol. 38, no. 2, pp. 114-129, 2005. https://doi.org/10.1016/j.jbi.2004.09.001

[15] A. Freitas and E. Curry, "Natural language queries over heterogeneous linked data graphs: A distributional-compositional semantics approach," in *Proceedings of the 19th International Conference on Intelligent User Interfaces*, 2014, pp. 279-288.

[16] R. Bendaoud, A. Napoli, and Y. Toussaint, "Formal concept analysis: A unified framework for building and refining ontologies," in *International Conference on Knowledge Engineering and Knowledge Management*, 2008: Springer, pp. 156-171.

[17] D. Diefenbach, V. Lopez, K. Singh, and P. Maret, "Core techniques of question answering systems over knowledge bases: A survey," *Knowledge and Information systems,* vol. 55, pp. 529-569, 2018. https://doi.org/10.1007/s10115-017-1100

[18] K. Vila and A. Ferrández, "Developing an ontology for improving question answering in the agricultural domain," in *Metadata and Semantic Research: Third International Conference, MTSR 2009, Milan, Italy, October 1-2, 2009. Proceedings 3*, 2009: Springer, pp. 245-256.

[19] Z. Y. Oğuz and Ç. Turhan, "Ontology-based semantic analysis of software requirements: A systematic mapping," in *2022 3rd International Informatics and Software Engineering Conference (IISEC)*, 2022: IEEE, pp. 1-6.

[20] M. D. Olvera-Lobo and J. Gutierrez-Artacho, "Language resources used in multi-lingual question-answering systems," *Online Information Review,* vol. 35, no. 4, pp. 543-557, 2011. https://doi.org/10.1108/14684521111161927

[21] G. Weikum, J. Graupmann, R. Schenkel, and M. Theobald, "Towards a statistically semantic web," in *International Conference on Conceptual Modeling*, 2004: Springer, pp. 3-17.

[22] E. Yudhono, S. Shaufiah, and I. L. Sardi, "Application of a question answering system in the ontology-based health domain and question templates to improve the quality of answers," *Journal La Multiapp,* vol. 5, no. 5, pp. 553-567, 2024. https://doi.org/10.37899/journallamultiapp.v5i5.1600

[23] R. Srinivasan and C. Subalalitha, "A thesaurus based semantic relation extraction for agricultural corpora," in *International Conference on Computational Intelligence in Data Science*, 2020: Springer, pp. 99-111.

[24] O. Danylovych, "The intensity of combinability of lexical-semantic groups of adjectives with lexical-semantic groups of nouns in the belle-lettres discourse of the modern English language," *Odessa Linguistic Journal,* vol. 13, pp. 1-10, 2019. https://doi.org/10.32837/2312-3192/13/1

[25] A. Reig Alamillo, D. Torres Moreno, E. Morales González, M. Toledo Acosta, A. Taroni, and J. Hermosillo Valadez, "The analysis of synonymy and antonymy in discourse relations: An interpretable modeling approach," *Computational Linguistics,* vol. 49, no. 2, pp. 429-464, 2023. https://doi.org/10.1162/coli_a_00477

[26] L. Bradeško *et al.*, "Contextualized question answering," *Journal of Computing and Information Technology,* vol. 18, no. 4, pp. 325-332, 2010. https://doi.org/10.2498/cit.1001912

[27] A. Rosso-Mateus, M. Montes-y-Gómez, P. Rosso, and F. A. González, "Deep fusion of multiple term-similarity measures for biomedical passage retrieval," *Journal of Intelligent & Fuzzy Systems,* vol. 39, no. 2, pp. 2239-2248, 2020. https://doi.org/10.3233/JIFS-179887

[28] M. Li, Y. Li, Y. Chen, and Y. Xu, "Batch recommendation of experts to questions in community-based question-answering with a sailfish optimizer," *Expert Systems with Applications*, vol. 169, p. 114484, 2021. https://doi.org/10.1016/j.eswa.2020.114484.

[29] M. Fiorelli, M. T. Pazienza, A. Stellato, and A. Turbati, "ART Lab infrastructure for semantic Big Data processing," in *Proceedings International Conference on High Performance Computing & Simulation (HPCS)*, 2014, pp. 327-334.

[30] P. Quaresma, V. Beires Nogueira, K. Raiyani, R. Bayot, and T. Gonçalves, *From textual information sources to linked data in the agatha project. lecture notes in computer science*. Cham: Springer, 2020.

[31] F. P. Romero, P. Julian-Iranzo, A. Soto, M. Ferreira-Satler, and J. A. Gallardo-Casero, *Fuzzy declarative approach for classifying unlabeled short texts using thesauri. lecture notes in computer science*. Berlin, Heidelberg: Springer, 2011.

[32] P. Julián and C. Rubio, "An efficient fuzzy unification method and its implementation into the Bousi~Prolog system," in *Proceedings of the FUZZ-IEEE 2010*, 2010, pp. 658–665.

[33] P. Velardi, R. Navigli, A. Cucchiarelli, and F. Neri, "Evaluation of OntoLearn, a methodology for automatic learning of domain ontologies," *Ontology Learning from Text: Methods, Evaluation, and Applications*, vol. 123, pp. 92–104, 2005.

[34] F. A. D. Tajouo, T. Noulamo, and J.-P. Lienou, "Procedure for the Contextual, Textual and Ontological Construction of Specialized Knowledge Bases," *European Journal of Electrical Engineering and Computer Science*, vol. 5, no. 1, pp. 62-67, 2021. https://doi.org/10.24018/ejece.2021.5.1.282.

[35] M. T. Pazienza, N. Scarpato, A. Stellato, and A. Turbati, "Semantic turkey: A browser-integrated environment for knowledge acquisition and management," *Semantic Web*, vol. 3, no. 3, pp. 279-292, 2012. https://doi.org/10.3233/SW-2011-0033

[36] M. Fiorelli, R. Gambella, M. T. Pazienza, A. Stellato, and A. Turbati, "Semi-automatic knowledge acquisition through CODA," in *Proceedings 27th International Conference on Industrial, Engineering Other Applications of Applied Intelligent Systems (IEA-AIE 2014)*, 2014.

[37] C. Caracciolo *et al.*, "The AGROVOC linked dataset," *Semantic Web*, vol. 4, no. 3, pp. 341-348, 2013. https://doi.org/10.3233/SW-13

[38] M. Fiorelli, M. T. Pazienza, A. Stellato, and A. Turbati, "CODA: Computer-aided ontology development architecture," *IBM Journal of Research and Development*, vol. 58, no. 2/3, pp. 1-14, 2014. https://doi.org/10.1147/JRD.2014.2307518

[39] D. Ferrucci and A. Lally, "UIMA: An architectural approach to unstructured information processing in the corporate research environment," *Natural Language Engineering*, vol. 10, no. 3-4, pp. 327-348, 2004. https://doi.org/10.1017/S1351324904003523