

Научная статья
УДК 004.33
doi: 10.17586/2713-1874-2025-3-83-95

МЕТОДИКА ОЦЕНКИ УЧЕБНЫХ КУРСОВ С ИСПОЛЬЗОВАНИЕМ МНОГОКОМПОНЕНТНЫХ МЕТРИК И ЭТАЛОННОГО СОПОСТАВЛЕНИЯ

Полина Анатольевна Шнайдер^{1✉}, *Илья Алексеевич Летенков*²,
*Максим Валерьевич Хлопотов*³

¹HelloFresh SE, Берлин, Германия

^{2,3}Университет ИТМО, Санкт-Петербург, Россия

¹beatrix.linkoln@gmail.com ✉, <https://orcid.org/0000-0002-4147-6561>

²ilya.letenkov@itmo.ru

³khlopotov@itmo.ru, <https://orcid.org/0000-0002-9053-027X>

Язык статьи – русский

Аннотация: В статье предложена система метрик для автоматизированной оценки университетских учебных курсов, основанная на анализе формулировок тем и их структурной организации. Целью исследования является разработка формализованной количественной методики оценки качества курсов, учитывающей как внутренние характеристики содержания, так и его соответствие эталонной дисциплине (при наличии). Методика включает эталон-независимые метрики, оценивающие логическую связность, сбалансированность и избыточность курса, а также эталон-зависимые метрики, позволяющие определить полноту охвата и тематическую релевантность. В качестве инструментов анализа применяются графовые представления, семантические эмбединги и метрики последовательности. Методика апробирована на корпусе дисциплин математического профиля. Результаты показали её эффективность в выявлении логических нарушений, пропусков ключевых тем и нерелевантного содержания. Разработанный подход может использоваться для поддержки проектирования и обновления учебных курсов, а также как основа автоматизированной обратной связи в цифровой образовательной среде. В качестве следующего этапа работы рассматривается интеграция системы с генеративными языковыми моделями и расширение набора метрик за счёт содержательных характеристик, таких как глубина раскрытия тем и уровни сложности.

Ключевые слова: анализ содержания, автоматизированная оценка, генерация курсов, графовое представление, структура курса, тематическая релевантность, формулировки тем, эталонная модель, эталон-независимые метрики, эталон-зависимые метрики

Ссылка для цитирования: Шнайдер П. А., Летенков И. А., Хлопотов М. В. Методика оценки учебных курсов с использованием многокомпонентных метрик и эталонного сопоставления // Экономика. Право. Инновации. 2025. Т. 13. № 3. С. 83–95. <http://dx.doi.org/10.17586/2713-1874-2025-3-83-95>.

AN APPROACH TO EVALUATING ACADEMIC COURSES THROUGH MULTI-FACTOR METRICS AND REFERENCE COURSE ALIGNMENT

Polina A. Shnaider^{1✉}, *Ilya A. Letenkov*², *Maksim V. Khlopotov*³

¹HelloFresh SE, Berlin, Germany

^{2,3}ITMO University, Saint Petersburg, Russia

¹beatrix.linkoln@gmail.com ✉, <https://orcid.org/0000-0002-4147-6561>

²ilya.letenkov@itmo.ru

³khlopotov@itmo.ru, <https://orcid.org/0000-0002-9053-027X>

Article in Russian

Abstract: This paper presents a system of metrics for the automated evaluation of university-level academic courses, based on the analysis of topic formulations and their structural organization. The objective of the study is to develop a formalized quantitative methodology for assessing course quality, accounting for both internal content characteristics and alignment with a reference syllabus (when available). The proposed approach includes reference-independent metrics that assess logical coherence, structural balance, and content redundancy, as well as reference-based metrics that evaluate topic coverage and thematic relevance. Analytical tools include graph-based representations, semantic embeddings, and sequence-based metrics. The methodology was tested on a corpus of mathematics-related university courses. The results demonstrate the system's effectiveness in identifying logical inconsistencies, missing core topics, and irrelevant content. The proposed approach can be used to support the design and revision of academic courses and to provide automated

feedback in digital learning environments. As a next step, the integration of the system with generative language models and the extension of the metric set to include content-level characteristics—such as topic depth and complexity levels—is being considered.

Keywords: automated evaluation, course generation, course structure, graph representation, reference-based metrics, reference-free metrics, semantic relevance, subject alignment, syllabus comparison, topic formulations

For citation Shnaider P. A., Letenkov I. A., Khlopotov M. V. An Approach to Evaluating Academic Courses Through Multi-Factor Metrics and Reference Course Alignment. *Ekonomika. Pravo. Innovacii*. 2025. Vol. 13. No. 3. pp. 83–95. (In Russ.). <http://dx.doi.org/10.17586/2713-1874-2025-3-83-95>.

Введение. Контроль качества содержания образовательных программ является ключевым фактором в обеспечении эффективности обучения, особенно в условиях перехода к цифровым образовательным средам. В университетской практике задачи, связанные с проектированием и оценкой учебных курсов, приобретают особую значимость на фоне расширения спектра дисциплин, развития междисциплинарных программ и необходимости быстрой адаптации содержания под запросы отрасли и академического сообщества.

В настоящее время оценка содержания курсов в университетах, как правило, осуществляется вручную на основе экспертных заключений, ведомственных стандартов и формальных требований (например, полноты рабочих программ и соответствия планируемым результатам обучения). Несмотря на свою значимость, такой подход не позволяет в полной мере выявлять логические несоответствия и оценивать информационную структуру курсов: дублирование тем, тематическая избыточность, слабая связанность и нерелевантный контент могут оставаться незамеченными. Особенно остро эта проблема проявляется в масштабных образовательных программах, включающих сотни курсов с разными форматами представления.

Актуальность автоматизированной оценки содержания курсов возрастает по мере внедрения цифровых инструментов: модульных редакторов, платформ коллективной разработки, систем поддержки учебных планов. В этих условиях университетам требуются формализованные количественные методы, способные обеспечивать масштабируемую, объективную и воспроизводимую оценку качества курсов как на этапе их проектирования, так и при последующем обновлении.

Литературный обзор. Необходимость разработки таких подходов подтверждается и

в современной научной литературе. В частности, Шалкина Т. Н. указывает на важность учёта не только полноты и корректности содержания электронных курсов, но и их соответствия дидактическим и структурным критериям [1]. В работе Старыгиной С. Д. рассматриваются автоматизированные модели оценки качества, позволяющие формализованно анализировать курсы на разных уровнях и получать количественные индикаторы их состояния [2]. Также существуют методики численной оценки качества онлайн-курсов, основанные на экспертной системе параметров и применимые в массовых образовательных проектах [3]. Дополнительный интерес представляют интеллектуальные методы оценки, использующие технологии анализа иерархий, онтологии и экспертные модели [4].

Наряду с классическими методиками всё большую популярность приобретают подходы, использующие инструменты машинного обучения и обработки естественного языка. Так, X. Chen и соавторы [5] предлагают модель оценки онлайн-курсов на основе анализа отзывов студентов с использованием тематического моделирования — в частности, модели латентного размещения Дирихле (LDA). Обнаруженные темы трактуются как критерии оценки, а итоговая иерархическая модель строится с помощью метода анализа иерархий (АНР). Хотя подход позволяет учитывать субъективные аспекты восприятия курса, он требует наличия массивов пользовательских данных и не затрагивает непосредственно структуру учебного материала.

Другой перспективный вектор связан с применением больших языковых моделей (LLM). В работе Wo Yuan и Jiazi Hu [6] рассматривается использование GPT-подобной модели для анализа содержательной полноты и семантической согласованности курсов. LLM используется как инструмент

экспертной поддержки для интерпретации тем, оценки релевантности и предварительного ранжирования. Однако такие модели функционируют как «чёрный ящик», что затрудняет интерпретируемость и верификацию результатов. На этом фоне предлагаемый в настоящей работе подход можно охарактеризовать как более детерминированный и интерпретируемый: он опирается на формализованные метрики, устойчив к вариативности формулировок и не требует ручной аннотации или дообучения моделей.

Таким образом, в научной и прикладной плоскости сохраняется потребность в разработке и валидации методов формализованной оценки качества учебных дисциплин, обеспечивающих как внутреннюю структурную проверку, так и сопоставление содержания с референсными моделями.

Целью настоящей работы является разработка и апробация системы метрик, позволяющей автоматизировано оценивать структуру, полноту и тематическую релевантность курсов.

Предлагаемый подход ориентирован на поддержку системного анализа образовательных данных и может использоваться в задачах контроля качества, валидации учебных планов, а также в системах рекомендаций и сопровождения образовательных траекторий.

Постановка задачи исследования. Целью настоящего исследования является разработка формализованной системы количественной оценки структуры и содержания учебных дисциплин, предназначенной для автоматизированной верификации качества курсов в цифровой образовательной среде.

В рамках достижения этой цели решаются следующие задачи.

- **Раздел: Введение в программирование**

- Основные понятия: алгоритм, программа, переменные
- Ввод и вывод данных
- Условные операторы и циклы

- **Раздел: Структуры данных**

- Массивы и списки
- Словари и множества
- Основы рекурсии

- 1) Определение ключевых характеристик учебных курсов, подлежащих оценке (структурная связность, полнота охвата, тематическая релевантность).

- 2) Построение набора эталон-зависимых и эталон-независимых метрик, отражающих указанные характеристики.

- 3) Экспериментальная валидация метрик на корпусе университетских дисциплин.

- 4) Анализ чувствительности и интерпретируемости метрик.

- 5) Демонстрация применимости подхода на примерах дисциплин различного профиля.

Структура учебной дисциплины. Учебная дисциплина моделируется как двухуровневая структура (рисунок 1), включающая:

- разделы – логически обособленные содержательные блоки (например, модули, тематические сегменты);

- темы – элементы содержания, сгруппированные внутри разделов. Такое представление обеспечивает формализованное описание курса, пригодное для анализа логики, сбалансированности и связности содержания. Хотя конкретные заголовки разделов и тем задаются преподавателем, на практике используются устойчивые словосочетания и типовые структуры, характерные для каждой предметной области. Например: в математических курсах: «Введение в линейную алгебру», «Методы решения СЛАУ», «Собственные значения и векторы»; в программировании: «Основы синтаксиса Python», «Работа с файлами», «Обработка исключений»; в экономике: «Рынки и конкуренция», «Основы налогообложения», «Финансовый анализ деятельности компании».

Рисунок 1 – Пример структуры курса

Источник: составлен авторами

Использование формулировок тем и разделов в качестве основной единицы анализа не претендует на отражение полной педагогической структуры дисциплины. Однако в условиях цифрового проектирования и типовой терминологии, сложившейся в предметных областях, такое приближённое представление позволяет проводить первичную структурную верификацию курса на этапе его проектирования или обновления.

Обоснование необходимости структурной оценки. Логически организованная структура способствует более глубокому и системному усвоению материала [7], снижает когнитивную нагрузку и повышает прозрачность траектории обучения.

Кроме того, она позволяет преподавателю и методисту выявлять дублирующиеся, нерелевантные или выпадающие из логики темы, что особенно актуально при проектировании или актуализации большого числа курсов.

Типы используемых метрик. В рамках предлагаемого подхода используются два класса метрик.

1) Эталон-независимые метрики – оценивают курс по внутренним признакам: логической связности тем внутри разделов, сбалансированности и степени повторяемости. Эти метрики не требуют сравнения с внешним эталоном и подходят для первичного анализа качества.

2) Эталон-зависимые метрики – сравнивают структуру курса с референсной дисциплиной и позволяют выявить пробелы в покрытии ключевых тем, избыточные элементы, отклонения в порядке и логике изложения.

В качестве эталона могут использоваться:

- дисциплины предыдущих лет;
- курсы родственного направления;
- референсные версии, разработанные внутри образовательной организации на основе накопленных данных и преподавательской практики.

Сравнение с эталоном позволяет обеспечить содержательную преемственность, проверять полноту и логическую стройность курса, а также повысить обоснованность принимаемых проектных решений в образовательной среде.

Методы и материалы исследования. Для формализованной оценки качества учебных дисциплин разработана система метрик, охватывающая три аспекта: структурную согласованность, тематическое покрытие и релевантность содержания. Часть метрик работает независимо от эталона, другая требует сравнения с референсной дисциплиной.

1) Структурные метрики. Позволяют оценить внутреннюю логику курса: согласованность тем, равномерность нагрузки и отсутствие дублирования.

Метрики:

- *semantic coherence* – связность тем внутри раздела,

- *structural balance* – равномерность распределения,

- *redundancy* – избыточность и повторы.

2) Метрики покрытия (*coverage*). Фиксируют, насколько содержание курса охватывает ключевые элементы эталона. Используются при проверке преемственности и актуальности.

Метрики:

- *semantic topic coverage* – темы эталона, представленные в курсе,

- *graph-based coverage* – охват содержательных связей.

3) Метрики релевантности. Выявляют нерелевантные или случайные темы, не соответствующие целям дисциплины.

Метрики:

- *relevance* – семантическое соответствие,

- *extra topics penalty* – доля нерелевантных тем.

Все метрики рассчитываются на основе формализованного представления курса в виде упорядоченной структуры: темы сгруппированы по разделам, порядок и логические связи сохраняются. Это позволяет проводить оценку как на уровне отдельных тем, так и на уровне макроструктуры курса в целом.

Структурные метрики. Структурные метрики предназначены для оценки внутренней логики и композиционной целостности учебной дисциплины. Они выявляют, насколько темы внутри и между разделами формируют связную, сбалансированную структуру.

Используются три подметрики: *semantic coherence*, *topic flow*, *structural balance*.

Метрики *semantic coherence* и *structural balance* являются эталон-независимыми – работают на основе структуры самого курса и полезны при проектировании новых дисциплин. Метрика *topic flow*, напротив, эталон-зависима и позволяет оценить логическую преемственность в сравнении с референсной структурой.

Метрики не стандартизируют содержание, а выявляют потенциальные структурные нарушения – дублирование, перегрузку, хаотичную последовательность. Все оценки опираются на формализованное представление курса в виде иерархии тем по разделам.

Semantic Coherence. Метрика *Semantic Coherence* оценивает тематическую согласованность тем внутри раздела. Предполагается, что темы, сгруппированные в один смысловой блок, должны быть логически связаны, но не идентичны: полное совпадение формулировок может свидетельствовать о дублировании, а избыточная разнородность – о нарушении логики курса.

Для расчёта используется модель *SentenceTransformer(all-MiniLM-L6-v2)*, основанная на мультязычной архитектуре *MiniLM*. Несмотря на англоязычную обучающую выборку, модель демонстрирует устойчивые результаты при анализе кратких русскоязычных формулировок благодаря универсальности векторных представлений. В качестве тем используются заголовки из структуры курса, без аннотаций или пояснений.

Каждая тема преобразуется в вектор размерности 384. Далее рассчитывается матрица косинусных сходств между темами в каждом разделе, и итоговая когерентность определяется по формуле:

$$\text{flow} = \left(1 - \frac{\text{inversions}}{\text{max_inversions}}\right) \cdot (1 - \text{penalty}_{\text{start}}), \quad (3)$$

где: *inversions* – количество инверсий между порядком тем в курсе и эталоне, *max_inversions* – максимально возможное число инверсий для данной длины последовательности, *penalty_{start}* – дополнительный штраф за то, что первая тема из эталона появляется слишком поздно в курсе.

Сравнение выполняется только по совпадающим темам. Темы, отсутствующие в эталоне, не учитываются напрямую – их влияние

$$\text{coherence}(R) = \bar{s}^2 \cdot (1 - 1.5 \cdot \sigma_s), \quad (1)$$

где \bar{s} – средняя косинусная схожесть между темами в разделе *R*, σ_s – стандартное отклонение этих значений (тематический разброс).

Метрика сбалансирована: она поощряет тематическую близость и штрафует за высокую дисперсию, однако не снижает значение при избыточной однородности. Для компенсации возможных завышений используется дополнительная метрика *Redundancy*, направленная на выявление повторяющихся формулировок.

Пример:

– Хорошо: «Алгоритмы сортировки» → «пузырьковая сортировка», «быстрая сортировка», «сортировка слиянием».

– Плохо: «Разное» → «SQL-запросы», «теория вероятностей», «ядро ОС».

Topic Flow. Метрика *Topic Flow* оценивает, насколько порядок тем в курсе соответствует логике эталонной структуры и поддерживает дидактическую траекторию. Она основана на модифицированной метрике *Kendall Tau Distance* – числа инверсий между совпадающими темами в двух упорядоченных последовательностях [8], с добавлением штрафа за позднее появление первой эталонной темы. Максимально возможное число инверсий рассчитывается как:

$$\text{max_inversions} = \frac{n(n-1)}{2}, \quad (2)$$

где *n* – число совпадающих тем. Таким образом, итоговая метрика принимает значения от 0 (полное несоответствие порядка) до 1 (полное совпадение):

отражают другие метрики, такие как *extra topics penalty* и *relevance*.

Примеры:

– Хорошо: логичная прогрессия — «Введение в Python» → «ООП» → «Асинхронное программирование».

– Плохо: порядок нарушен — «Базы данных» → «Циклы» → «Форматы сериализации», при этом базовые темы появляются позже продвинутых.

Метрика дополняет показатели покрытия (coverage) и фиксирует отклонения от дидактически выверенного порядка, что особенно важно при проектировании новых курсов и валидации образовательных траекторий.

Structural Balance. Метрика Structural Balance оценивает равномерность распределения тем между разделами. Её задача – выявление перекосов: перегрузки одних разделов при недостаточности других, что может затруднять восприятие материала.

$$\text{balance} = 1 - \frac{\frac{\sigma_n + \frac{\Delta n}{\max(n)}}{\bar{n}}}{2} - \delta, \quad (4)$$

где σ_n – стандартное отклонение числа тем в разделах, \bar{n} – среднее количество тем, Δn – максимальный скачок между соседними разделами, δ – дополнительный штраф за чрезмерную перегрузку одного из разделов.

Метрика штрафует как значительные диспропорции в числе тем, так и резкие переходы между разделами. При этом она не требует абсолютной симметрии и не интерпретирует различия в содержании как методическую ошибку.

Метрика не требует полной симметрии и не оценивает учебную трудоёмкость тем. В рамках одной дисциплины возможны как крупные модули, так и компактные тематические блоки. Поэтому высокая

сбалансированность по данной метрике – не гарантия качества, а сигнальный индикатор, требующий экспертной интерпретации.

Примеры:

– Хорошо: каждый из 5 разделов содержит по 3–5 тем.

– Плохо: один раздел содержит 15 тем, два других – по одной.

Метрика используется как вспомогательный инструмент на этапе проектирования, позволяя выявлять потенциальную структурную асимметрию, но не подменяет содержательный анализ курса.

Метрики покрытия. Метрики покрытия (coverage metrics) позволяют оценить, насколько содержание курса охватывает ключевые темы эталона. Они не требуют полного совпадения формулировок или структуры, но фиксируют пробелы и отклонения при проектировании или обновлении дисциплин. Используются два подхода: семантическое покрытие и графовое покрытие связей.

Semantic Topic Coverage. Оценивает, какие темы эталона представлены в курсе с точки зрения смысла (без учёта порядка). Каждая тема эталона сравнивается с темами курса по косинусному сходству эмбедингов (SentenceTransformer all-MiniLM-L6-v2), и считается совпавшей, если превышен порог θ , например, 0.8 (формула 5):

$$\text{coverage} = \frac{|\{t \in T_{\text{ref}} | \exists s \in T_{\text{course}} : \cos(\text{emb}(t), \text{emb}(s)) > \theta\}|}{|T_{\text{ref}}|}, \quad (5)$$

где T_{ref} – множество тем эталона, T_{course} – множество тем анализируемого курса, θ – порог семантической близости.

Примеры:

– Хорошо: курс охватывает все ключевые темы эталона, пусть и в изменённой формулировке: «Основы алгоритмизации», «Типы данных», «Условные конструкции» → «Алгоритмическое мышление», «Базовые структуры данных», «Операторы выбора».

– Плохо: в курсе отсутствуют несколько центральных тем эталона, например, «Рекурсия», «Сложность алгоритмов», что указывает на неполный охват.

Метрика особенно полезна для дисциплин с гибкой структурой и на ранних этапах проектирования.

Graph-Based Structural Coverage. Дополняет предыдущую метрику, оценивая сохранение связей между темами. Эталон представляется как граф:

вершины – темы,

рёбра – логические связи (обычно последовательности тем внутри разделов).

Темы сопоставляются с вершинами эталона (порог близости – 0.75), затем строится частичный граф и оценивается покрытие по формуле (6):

$$\text{coverage} = \frac{|V_{\text{match}}| + |E_{\text{match}}|}{|V_{\text{ref}}| + |E_{\text{ref}}|}, \quad (6)$$

где V_{match} , E_{match} – число совпавших тем и рёбер, V_{ref} , E_{ref} – общее число тем и рёбер в эталонном графе.

Метрика чувствительна к нарушению дидактической логики и полезна для задач, связанных с обновлением дисциплин, валидацией учебных траекторий и обеспечением содержательной преемственности.

Примеры:

– Хорошо: курс охватывает почти все темы эталона, включая как центральные, так и периферийные узлы графа. Даже при отличиях в формулировках и структуре сохраняются ключевые связи между темами – «Типы данных» → «Алгоритмы поиска» → «Сортировки» → «Анализ сложности».

– Плохо: курс включает лишь отдельные, изолированные темы, не соединённые в логическую цепочку. Например, есть «Введение» и «Современные подходы», но полностью отсутствуют «Базовые

алгоритмы» и «Структуры данных», необходимые для понимания более сложных элементов.

Метрику *Graph-Based Structural Coverage* целесообразно применять, когда важно оценить не только наличие тем, но и сохранность логических связей между ними. Это особенно актуально для дисциплин, где последовательность тем влияет на понимание, например, в курсе «Введение в искусственный интеллект», где освоение методов должно опираться на ключевые зависимости между понятиями.

Redundancy. Оценивает наличие избыточных тем, семантически или лексически схожих. Рассчитывается на основе сходства между парами тем и коррекции по масштабам курса (формула 7):

$$\text{redundancy} = \min \left(1.0, \bar{p} \cdot \left(\frac{n_{\text{topics}}}{n_{\text{pairs}}} \right)^\alpha + \beta \cdot n_{\text{lex_dupl}} \right), \quad (7)$$

где \bar{p} – средний штраф по семантически близким парам, n_{pairs} – число оценённых пар, n_{topics} – общее число тем в курсе, $n_{\text{lex_dupl}}$ – количество лексических дубликатов, α, β – коэффициенты коррекции (обычно подбираются из диапазонов $0.5 \leq \alpha \leq 1, 0.05 \leq \beta \leq 0.2$).

Примеры:

– Хорошо: все темы курса уникальны по содержанию, даже при частичном совпадении терминов («Машинное обучение» vs «Глубокое обучение»).

– Плохо: в курсе присутствуют темы «SQL-запросы», «Язык SQL», «Основы SQL» — при этом они расположены в разных разделах и слабо различаются по сути.

Метрика *Redundancy* относится к эталонно-независимым и может использоваться на любом этапе проектирования дисциплины. Осо-

бенно эффективно применять её в сочетании с *semantic coherence* и *structural balance* для анализа внутренней согласованности курса.

Метрики релевантности. Метрики данной группы позволяют определить, насколько темы курса соответствуют тематическому фокусу эталона. Они ориентированы на выявление нерелевантных или избыточных тем, не отражённых в предметной области дисциплины.

Relevance. Метрика *Relevance* вычисляет долю тем курса, для которых существует семантически близкая тема в эталоне. Для каждой темы из курса определяется максимальное значение косинусной близости к любой теме эталона (по эмбедингам *Sentence Transformer* ("all-MiniLM-L6-v2")). Если оно превышает пороговое значение (например, 0.7), тема считается релевантной (формула 8):

$$\text{relevance} = \frac{|\{t \in T_{\text{course}} \mid \max_{x \in T_{\text{ref}}} \cos(\text{emb}(t), \text{emb}(x)) > \theta\}|}{|T_{\text{course}}|}, \quad (8)$$

где T_{course} – множество тем анализируемого курса, T_{ref} – множество тем эталонной дисциплины, $\text{emb}(x)$ – эмбединг формулировки темы x , $\cos(\cdot, \cdot)$ – косинусное сходство между двумя эмбедингами, θ – порог семантической близости (например, 0.7), числитель – число тем курса, для которых найдена как минимум одна тема в эталоне с

семантической близостью выше порога, знаменатель – общее число тем в курсе.

Метрика оценивает каждую тему индивидуально, отражая, насколько она вписывается в семантическое пространство эталонной дисциплины.

Примеры:

– Хорошо: Эталон по дисциплине «Ана-

лиз данных» содержит темы: «регрессия», «кластеризация», «оценка моделей». Курс включает: «линейная регрессия», «иерархическая кластеризация», «метрики качества» → Темы не совпадают дословно, но семантически близки, релевантность высокая.

– Плохо: Эталон: «Алгоритмы и структуры данных» Курс включает темы: «основы

менеджмента», «документооборот», «эмоциональное выгорание» → Темы не связаны с исходной предметной областью, семантическое пересечение отсутствует — релевантность близка к нулю.

Extra Topics Penalty. Дополняющая метрика, фиксирующая долю тем, не имеющих семантически близких соответствий (формула 9):

$$extra_topics_penalty = \frac{|\{t \in T_{course} | \max_{x \in T_{ref}} \cos(emb(t), emb(r)) < \theta\}|}{|T_{course}|}, \quad (9)$$

Она использует те же обозначения, что и Relevance, и отличается только знаком сравнения в числителе с параметром θ . Таким образом, обе метрики арифметически дополняют друг друга до единицы при одинаковом пороге.

Примеры:

– Хорошо: Эталон: «Введение в машинное обучение» Курс: «регрессия», «деревья решений», «регуляризация», «кросс-валидация» → Все темы находят семантически близкие аналоги. → Extra Topics Penalty ≈ 0.0 .

– Плохо: Эталон: «Программирование на Python» Курс: «мотивация в обучении», «инфлюенс-маркетинг», «устойчивое развитие» → Темы полностью нерелевантны. → Extra Topics Penalty ≈ 1.0 .

Пограничный случай: Эталон: «Базы данных», Курс из 8 тем, из которых 6 совпадают, а 2 отсутствуют в эталоне: «создание резюме», «развитие soft skills» → Extra Topics Penalty = $2 / 8 = 0.25$.

Формирование итоговой оценки. Для получения интегральной количественной оценки качества курса используется агрегированная модель, объединяющая значения ключевых метрик из трёх категорий: структурных, метрик покрытия и метрик релевантности. Каждая категория отражает отдельный аспект качества: логическую целостность, полноту охвата и тематическое соответствие содержанию эталонной дисциплины.

Итоговая оценка вычисляется по формуле (10):

$$FinalScore = w_s \cdot S + w_c \cdot C + w_r \cdot R, \quad (10)$$

где S – агрегированное значение структурных метрик (semantic coherence, topic flow,

structural balance, redundancy); C – агрегированное значение метрик покрытия (semantic topic coverage, graph-based coverage); R – показатель релевантности (Relevance или $1 - ExtraTopicsPenalty$); w_s , w_c , w_r – веса, определяющие вклад каждой группы метрик (по умолчанию: $w_s = 0,33$, $w_c = 0,33$, $w_r = 0,34$).

Учет наличия эталона. Наличие эталонной дисциплины влияет на доступность отдельных групп метрик:

Если эталон доступен, используются все три категории показателей.

Если эталон отсутствует, итоговая оценка формируется только на основе эталон-независимых метрик: semantic coherence, structural balance, redundancy.

В этом случае:

$$FinalScore = \frac{1}{n} \sum_{i=1}^n M_i \quad (11)$$

где M_i – значения доступных метрик, n – их число.

Рекомендуемые веса и адаптация. По умолчанию веса w_s , w_c , w_r устанавливаются равными, что обеспечивает сбалансированный учёт структуры, покрытия и релевантности: $w_s = 0,33$, $w_c = 0,33$, $w_r = 0,34$.

Однако в зависимости от предметной области и цели анализа веса могут адаптироваться: для технических и фундаментальных дисциплин (математика, программирование) может усиливаться вклад структурных характеристик: $w_s \in [0,35; 0,5]$ для междисциплинарных или вариативных курсов предпочтение может отдаваться семантической релевантности и охвату: $w_r \in [0,4; 0,6]$, $w_c \in [0,2; 0,4]$.

Результаты. Описанная система метрик была применена для автоматизирован-

ной оценки университетских дисциплин, относящихся к направлению «Математика». Для формирования выборки использовались все курсы из цифровой базы, названия которых содержали ключевые слова, связанные с математикой. Это позволило выделить корпус дисциплин, близких по тематике, но отличающихся по структуре, уровню детализации и форме подачи материала.

В рамках эксперимента использовался единый эталон, представляющий собой наиболее полное, логически связное и терминологически насыщенное описание базового математического курса. Эталон был автоматически выбран на основе количества тем, охватывающих ключевые понятия предметной области. Учитывая широкую вариативность математических дисциплин, такой подход может быть применён исключительно

для первичной автоматизированной диагностики.

В перспективе может быть реализована схема сопоставления курсов с несколькими эталонами или автоматическая кластеризация по содержательной близости.

На рисунке 2 представлен пример сопоставления двух курсов, приближённых по тематике, но различающихся по уровню структурной согласованности:

В первом случае (итоговая оценка: 0.933) зафиксировано полное покрытие ключевых тем, педагогически логичная структура и внутренняя согласованность. Во втором случае (оценка: 0.115) наблюдаются тематическая фрагментированность, перегруженные или отсутствующие разделы и преобладание нерелевантного эталону содержания.

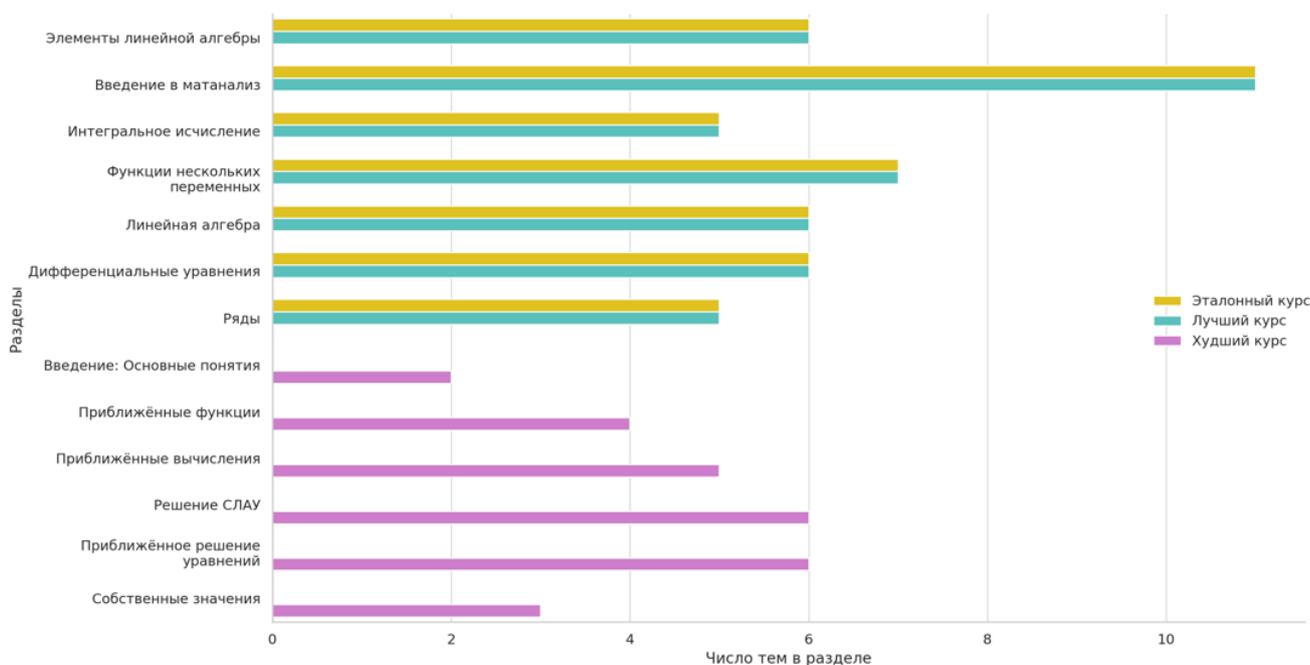


Рисунок 2 – Распределение тем по разделам в эталонном, «лучшем» и «худшем» курсе по направлению «математика»

Источник: составлен авторами

Дополнительно была выполнена массовая автоматизированная оценка всех курсов математического профиля, представленных в системе. Рисунок 3 отражает распределение значений метрик по всей выборке. Особенно выделяется смещение в сторону низких значений по *topic flow*, *semantic topic coverage* и *graph-based coverage*, что указывает на частые отклонения от эталонной структуры и недостаточное покрытие ключевых понятий.

Такая картина является ожидаемой, поскольку для всех курсов использовался единый эталон – обобщённая версия базового математического курса. Это неизбежно приводит к снижению показателей у дисциплин, отличающихся по направленности, уровню или методике преподавания. Но даже при этом условии модель позволяет выявлять курсы с критическими отклонениями по структуре, тематике или релевантности и может служить инструментом первичной диагностики.

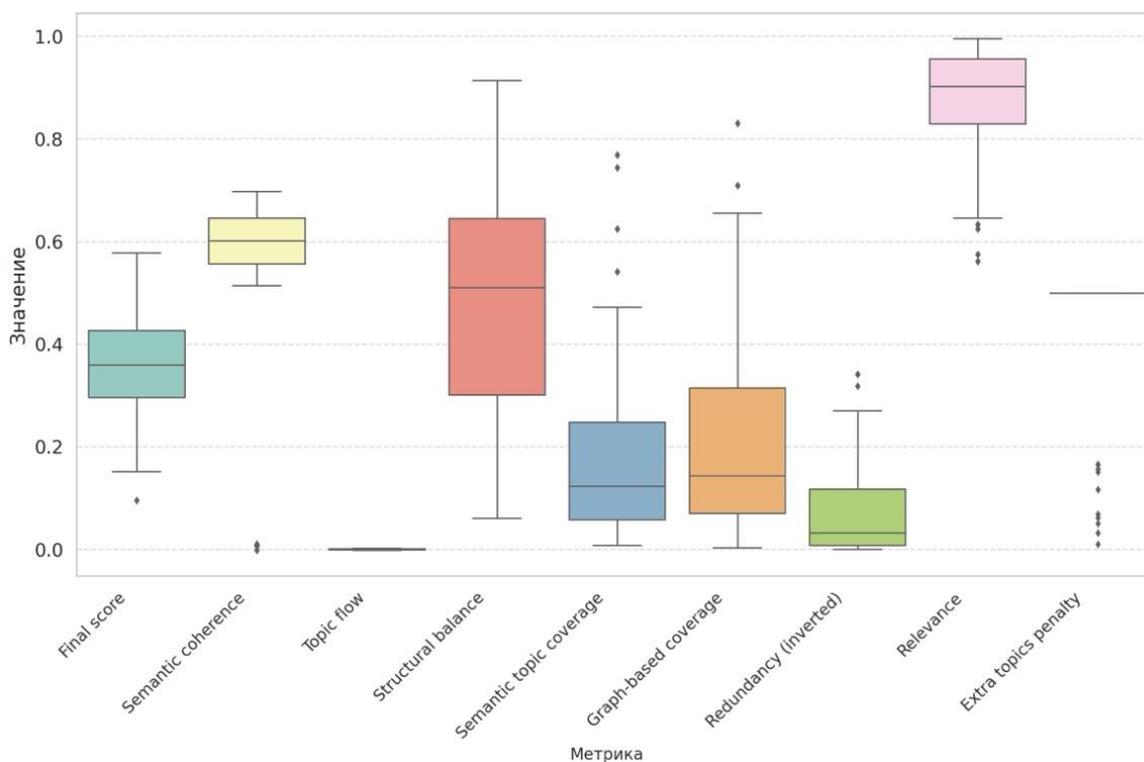


Рисунок 3 – Распределение значений метрик по курсам с названием «математика»
 Источник: составлен авторами

Система также была протестирована на курсе «Иностранный язык в профессиональной деятельности», представленного в нескольких версиях.

1) Оценка без эталона (структурная проверка).

В первой версии курс был проанализирован без сопоставления с эталоном, только с использованием эталон-независимых метрик.

Результаты показали:

- высокий показатель structural_balance (1.000),
- умеренную semantic_coherence (0.559),
- отсутствие избыточности (redundancy = 0.000).

Вывод: курс хорошо структурирован, темы логично сгруппированы и сбалансированы. (см. рисунок 4).

2) Оценка с эталоном.

При подключении референсной дисциплины (ID: 36517) были активированы метрики покрытия и релевантности (рисунок 5). Курс продемонстрировал:

- высокое покрытие тем (sequence_coverage = 1.000),
- полную релевантность (relevance = 1.000),

– несколько низкие значения по topic_flow (0.267) и graph-based coverage (0.545).

Вывод: несмотря на высокое тематическое соответствие, структура изложения отличается от эталона.

Интерпретация различий. На рисунке 6 показано, что курс и эталон содержат сопоставимый набор тем, однако порядок их изложения отличается. В большинстве дисциплин такое расхождение может повлиять на логичность подачи и усвояемость материала. Однако в случае «Иностранного языка в профессиональной деятельности» это не критично по следующим причинам:

- курс ориентирован на применение языка в типовых профессиональных ситуациях, а не на последовательное освоение понятий с нарастающей сложностью;
- каждая тема является относительно автономным практическим модулем (например, «Самопрезентация», «Составление документации», «Чтение технических инструкций»), и может осваиваться независимо от других;
- преподавание строится не на логике теоретической прогрессии, а на гибкой

адаптации под цели группы и текущие задачи студентов.

Вывод: низкие значения по topic flow и graph-based coverage отражают отклонение от структуры эталона, но не свидетельствуют о нарушении качества. Для таких дисциплин более важны показатели semantic coherence, relevance и semantic topic coverage, которые в данной реализации достигают высоких значений.

Таким образом, система может использо-

ваться как в задачах внутреннего контроля качества, так и для автоматизированного сопоставления и обновления дисциплин, поддержания преемственности и унификации содержания в рамках образовательных программ. Она позволяет автоматически выявлять курсы:

- с низкой тематической релевантностью,
- с избыточным или недостаточным покрытием,
- с нарушенной структурной логикой.

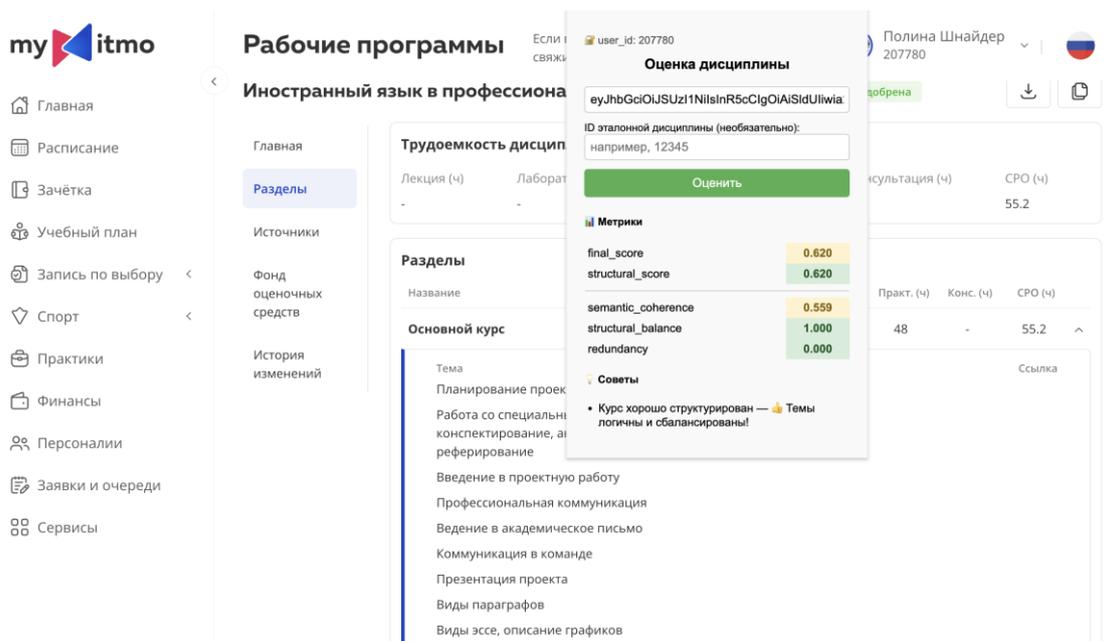


Рисунок 4 – Оценка дисциплины «Иностранный язык в профессиональной деятельности» без использования эталона

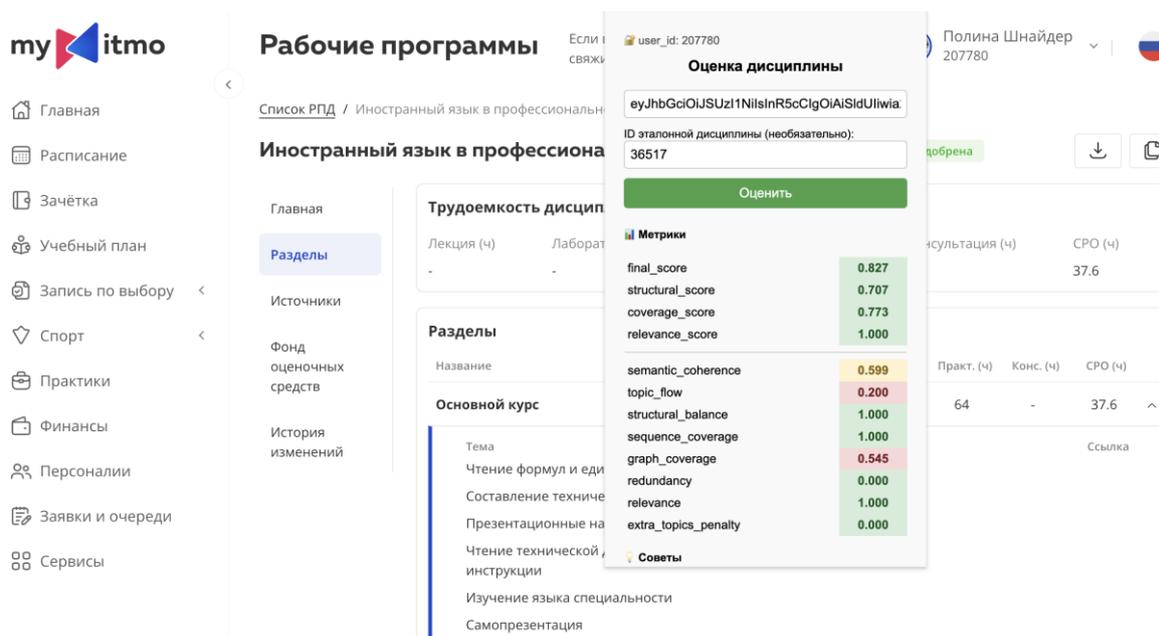


Рисунок 5 – Оценка дисциплины «Иностранный язык в профессиональной деятельности» с использованием эталона

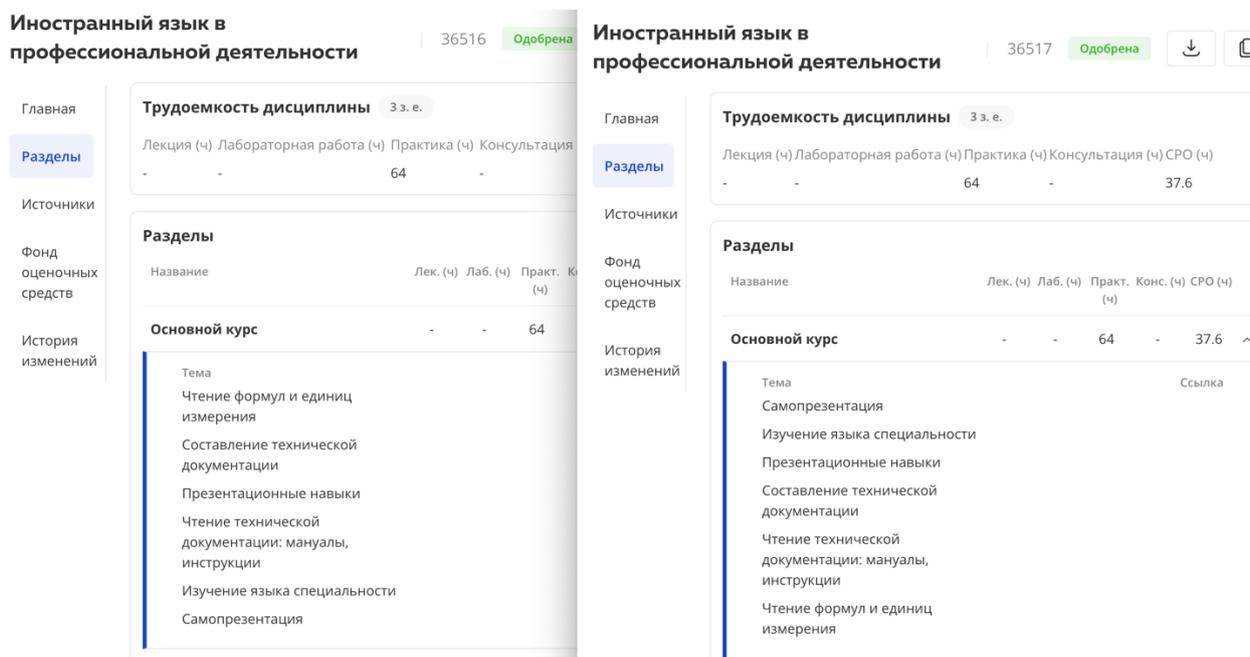


Рисунок 6 – Сопоставление структур сравниваемых дисциплин

Выводы и рекомендации. Предложенная система метрик обеспечивает автоматизированную количественную оценку учебных курсов на основе их структурной логики, полноты содержания и тематической релевантности по отношению к эталонной дисциплине. Подход продемонстрировал способность выявлять дисциплины с пропущенными ключевыми темами, избыточностью, фрагментированной структурой или отклонениями от педагогически оправданного порядка изложения.

Ключевыми преимуществами метода являются:

- масштабируемость – возможность оценки десятков и сотен курсов без участия экспертов;
- семантическая устойчивость – благодаря использованию эмбедингов, метод сглаживает различия в формулировках;
- воспроизводимость – оценки можно сравнивать между дисциплинами и версиями.

Вместе с тем подход имеет ряд ограничений:

- он ориентирован на структурные признаки, не учитывая глубину раскрытия тем, используемые методики преподавания и уровень подготовки студентов;
- полученные оценки не являются нормативными, а служат ориентиром: отклонения

от эталона могут быть осознанными и педагогически оправданными.

Таким образом, система не подменяет экспертную оценку, а выступает как инструмент аналитической поддержки преподавателей, методистов и администраторов.

Перспективы развития.

1) Интеграция с генеративными языковыми моделями для генерации структуры курсов и самопроверки моделей при создании содержания [9].

2) Формирование автоматизированной обратной связи преподавателю на основе сильных и слабых сторон структуры курса.

3) Расширение системы за счёт содержательных метрик, таких как глубина раскрытия, наличие оценочных средств, уровни когнитивной сложности.

4) Анализ изменений структуры во времени для отслеживания эволюции курсов и её влияния на качество.

5) Поддержка сопоставления и унификации дисциплин в рамках образовательных программ и сетевого взаимодействия вузов.

Таким образом, система может использоваться не только как средство контроля качества, но и как инструмент анализа, планирования и сопровождения учебных дисциплин, особенно в условиях цифровой трансформации образования.

Список источников

1. Шалкина Т. Н. Модель комплексной оценки качества электронных образовательных изданий и ресурсов // Педагогическая информатика. – 2013. – № 4. – С. 85–91.
2. Старыгина С. Д., Нуриев Н. К., Гарифьянов Н. Ф. Оценка качества объекта (учебного курса): построение модели, автоматизация расчётов, примеры реализации // Образовательные технологии и общество. – 2018. – Т. 21. – № 2. – С. 390–405.
3. Старыгина С. Д., Нуриев Н. К., Нургалиева А. А. Численная оценка качества учебных on-line курсов // Информатика и образование. – 2018. – № 7. – С. 68–76.
4. Богданова А. В., Глазова В. Ф., Коновалова Е. Ю. Интеллектуальные технологии оценки качества дистанционных учебных курсов в высшем образовании // Балтийский гуманитарный журнал. – 2017. – Т. 6. – № 1(18). – С. 82–86.
5. Chen X., Xie H., Tao X., Wang F. L., Cao J. Leveraging Text Mining and Analytic Hierarchy Process for the Automatic Evaluation of Online Courses // International Journal of Machine Learning and Cybernetics. – 2024. – Т. 15. – № 11. – С. 4973–4998. – DOI:10.1007/s13042-024-02203-6. – Текст: электронный. (In Eng.).
6. Yuan B., Hu J. An Exploration of Higher Education Course Evaluation by Large Language Models. – 2024. – DOI: 10.48550/arXiv.2411.02455. – Текст: электронный. (In Eng.).
7. Миннегалиева Ч. Б. Некоторые проблемы применения дистанционных образовательных технологий // Информационные технологии и системы. – 2013. – № 1 (70). – С. 54–58.
8. Erbisti B., Marzagão D. K., Braganholo V. An Extension to Kendall's Tau Metric to Evaluate Dissimilarities Between Data Series // Journal of the Brazilian Computer Society. – 2024. – Т. 30. – № 1. – URL: <https://journals-sol.sbc.org.br/index.php/jbcs/article/view/2803> (дата обращения 20.06.2025). – Текст: электронный. (In Eng.).
9. Wang S., Xu T., Li H., Zhang C., Liang J., Tang J., Yu P.S., Wen Q. Large Language Models for Education: A Survey and Outlook // arXiv preprint arXiv:2403.18105. – 2024. – URL: <https://arxiv.org/abs/2403.18105> (дата обращения 20.06.2025). – Текст: электронный. (In Eng.).

References

1. Shalkina T. N. A Model for Comprehensive Evaluation of the Quality of Electronic Educational Resources. *Pedagogicheskaya informatika*. 2013. No. 4. pp. 85–91. (In Russ.).
2. Sarygina S. D., Nuriev N. K., Garifyanov N. F. Evaluation of Course Quality: Model Development, Automation of Calculations, Implementation Examples. *Obrazovatel'nye tekhnologii i obshchestvo*. 2018. Vol. 21. No. 2. pp. 390–405. (In Russ.).
3. Sarygina S. D., Nuriev N. K., Nurgalieva A. A. Quantitative Assessment of Online Course Quality. *Informatika i Obrazovanie*. 2018. No. 7. pp. 68–76. (In Russ.).
4. Bogdanova A. V., Glazova V. F., Konovalova E. Yu. Intelligent Technologies for Evaluating the Quality of Distance Learning Courses in Higher Education. *Baltiiskii gumanitarnyi zhurnal*. 2017. Vol. 6. No. 1 (18). pp. 82–86. (In Russ.).
5. Chen X., Xie H., Tao X., Wang F. L., Cao J. Leveraging Text Mining and Analytic Hierarchy Process for the Automatic Evaluation of Online Courses. *International Journal of Machine Learning and Cybernetics*. 2024. Vol. 15. No. 11. pp. 4973–4998. DOI:10.1007/s13042-024-02203-6.
6. Yuan B., Hu J. An Exploration of Higher Education Course Evaluation by Large Language Models. 2024. DOI: 10.48550/arXiv.2411.02455.
7. Minnegalieva Ch. B. Some Problems of Using Distance Learning Technologies. *Informatsionnye tekhnologii i sistemy*. 2013. No. 1 (70). pp. 54–58. (In Russ.).
8. Erbisti B., Marzagão D.K., Braganholo V. An Extension to Kendall's Tau Metric to Evaluate Dissimilarities Between Data Series. *Journal of the Brazilian Computer Society*. 2024. Vol. 30. No. 1. Available at: <https://journals-sol.sbc.org.br/index.php/jbcs/article/view/2803>
9. Wang S., Xu T., Li H., Zhang C., Liang J., Tang J., Yu P.S., Wen Q. Large Language Models for Education: A Survey and Outlook. *arXiv preprint arXiv:2403.18105*. 2024. Available at: <https://arxiv.org/abs/2403.18105>