DOI: 10.58168/MIST2026_511-529
УДК 004.9

ОНЛАЙН-ПЛАТФОРМА ДЛЯ ПОДГОТОВКИ СТУДЕНТОВ СПО К ВСЕРОССИЙСКОЙ ПРОВЕРОЧНОЙ РАБОТЕ (ВПР) С ИСПОЛЬЗОВАНИЕМ ОБУЧАЮЩЕЙ ИИ МОДЕЛИ

Т.В. Скворцова[1], П.А. Пискурский[1], О.А. Авилова[1]

[1]ФГБОУ ВО «Воронежский государственный лесотехнический университет имени Г.Ф. Морозова»

Аннотация. В работе рассматривается разработка веб-платформы для подготовки студентов среднего профессионального образования к всероссийской проверочной работе с применением искусственного интеллекта для автоматической оценки развернутых ответов. В основе решения лежит гибридный подход, объединяющий латентно-семантический анализ, косинусное сходство и дообучаемые трансформерные модели. Платформа позволяет сократить время проверки ответов преподавателями на 60–80%, снизить количество пересдач и повысить объективность оценивания. Решение разработано с учётом требований к хранению данных на территории Российской Федерации и совместимости с отечественными образовательными системами.

Ключевые слова: автоматическая оценка ответов, искусственный интеллект в образовании, всероссийская проверочная работа, среднее профессиональное образование, латентно-семантический анализ, гибридные NLP-модели, цифровая образовательная среда.

ONLINE PLATFORM FOR PREPARING STUDENTS OF SECONDARY EDUCATION FOR THE ALL-RUSSIAN TEST (VPR) USING A TRAINING AI

T.V. Skvortsova[1], P.A. Piskursky[1], O.A. Avilova[1]

[1]Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. This paper examines the development of a web platform for preparing secondary vocational education students for the All-Russian assessment using artificial intelligence for the automatic evaluation of detailed answers. The solution is based on a hybrid approach combining latent semantic analysis, cosine similarity, and retrainable transformer models. The platform reduces the time teachers spend checking answers by 60–80%, reduces the number of retakes, and improves the

objectivity of assessment. The solution is designed to meet data storage requirements within the Russian Federation and ensure compatibility with domestic educational systems.

Keywords: automatic assessment, artificial intelligence in education, All-Russian test, secondary vocational education, latent semantic analysis, hybrid NLP models, digital educational environment.

Введение

Подготовка студентов среднего профессионального образования (СПО) к всероссийской проверочной работе (ВПР) остаётся сложной и ресурсоёмкой задачей, в значительной части из-за ручной проверки развернутых ответов, задержек обратной связи и ограниченности специализированных цифровых инструментов. На уровне системы образования проблема подтверждается регулярными аналитическими отчётами и обзорами: высокая нагрузка на преподавателей, дефицит объективной и быстрой оценки, нерациональные расходы времени и средств [1, 2]. Дополнительно обсуждается социально-профессиональный аспект – стресс и выгорание педагогов, что усиливается циклом тренировочных попыток перед ВПР и ростом пересдач [3]. В практической плоскости значимо и экологическое измерение: массовые распечатки контрольно-измерительных материалов и тренировочных заданий приводят к ощутимым расходам древесины и сопутствующих материалов, особенно при многократных попытках подготовки.

Существующие популярные решения в России фокусируются либо на типовых тестах и банках задач, либо на универсальных системах управления обучением (СУО), не предоставляя надёжной автоматической семантической оценки развернутых ответов. Так, «Сдам ГИА» ориентирован преимущественно на тесты с выбором варианта и не выполняет объективную автоматическую проверку свободных текстовых ответов [7], а Moodle в стандартной конфигурации предполагает ручную проверку эссе и развернутых ответов преподавателем [8]. Итогом становится замедленная обратная связь, субъективность и ограниченная аналитика по темам и навыкам.

В этой работе предлагается и исследуется научно-технический подход и прототип онлайн-платформы, специализирующейся на подготовке к ВПР в СПО и обеспечивающей автоматическую оценку развернутых ответов на основе латентно-семантического анализа (ЛСА) и косинусного сходства. Выбор ЛСА и близкостных метрик опирается на их фундаментальную обоснованность в информационном поиске и семантическом сопоставлении текстов [5, 6], а также на практическую воспроизводимость в условиях ограниченных вычислительных

ресурсов и гибкость настройки под рубрики ВПР. Для проверки гипотезы и получения воспроизводимых результатов используется открытый публичный датасет коротких ответов с экспертными оценками (ASAP – Автоматизированная оценка коротких ответов, платформа Kaggle) [9]. Несмотря на англоязычное происхождение датасета, его структура (пары «ответ–оценка», рубрикатор, балльная шкала) методологически релевантна задачам СПО/ВПР, а часть экспериментов сопровождается переводной адаптацией для оценки переносимости подхода на русскоязычные материалы.

Цели и новизна исследования состоят в том, чтобы:

– обосновать применимость ЛСА и косинусного сходства для автоматической семантической проверки развернутых учебных ответов в условиях ВПР-конкурентных рубрик;

– показать воспроизводимость результатов на публичном наборе данных, дать метрики точности и согласованности с экспертной оценкой, а также описать полный конвейер предварительной обработки;

– продемонстрировать практическую ценность для образовательной организации через оценочные расчёты экономического эффекта, уместность интеграции в цифровую среду и снижение нагрузки преподавателей.

Научная новизна работы – в систематическом объединении классического ЛСА-подхода с рубричной моделью оценивания ВПР и реализацией объяснимой проверки свободного ответа через семантическую близость к эталонным фрагментам рубрик. В отличие от существующих банков задач и универсальных СУО, исследуемый подход стремится обеспечить: ускорение циклов обратной связи, унификацию критериев, корреляцию с экспертными баллами и в перспективе снижение потребности в бумажных материалах за счёт онлайн-режима.

Фокус работы – не на замене преподавателя, а на уменьшении рутины и повышении качества данных. Платформа и описанные методы ориентированы на интеграцию в российскую образовательную инфраструктуру и на соблюдение требований хранения данных, что важно для реального внедрения в СПО. Предложенный конвейер учитывает этапы нормализации текста, лемматизации, построения взвешенных представлений (ТФ-ИДФ с последующим сингулярным разложением), калибровку порогов и сопоставление с эталонными рубриками ВПР [5, 6], что далее детализируется в методическом разделе.

Ключевые вклад и практические результаты, которых мы стремимся достигнуть в статье:

– верификация базовой точности ЛСА+косинус (точность, полнота, F1, доля совпадений с экспертом в диапазоне ±1 балла) на публичном датасете [9];

– демонстрация объяснимости оценок (подсветка покрытых рубричных элементов, коэффициенты близости по каждому компоненту ответа);

– анализ ограничений, типов ошибок и условий воспроизводимости (разметка данных, объёмы, влияние предварительной обработки).

Обзор проблем и контекста

Проблематика подготовки студентов СПО к ВПР складывается из взаимосвязанных управленческих, педагогических, технологических и экологических факторов. Нагрузку на преподавателей, необходимость объективности оценивания, дефицит инструментов автоматической семантической проверки развернутых ответов и бумажные издержки в массовом формате подготовки подтверждают как отраслевые обзоры, так и практические наблюдения в учреждениях образования [1, 2, 3]. При этом действующие инициативы цифровизации (федеральный проект «Цифровая образовательная среда») акцентируют доступность и инфраструктуру, но не решают специфическую проблему автоматической оценки свободного текста по рубрикам ВПР [11].

Во-первых, высокая нагрузка на преподавателей при проверке развернутых ответов приводит к задержкам обратной связи, росту пересдач и апелляций, и в конечном счёте снижению мотивации и качества подготовки. В аналитических материалах по ВПР и в публичных обсуждениях фиксируется дефицит времени на проверку, необходимость унификации критериев, а также стресс и выгорание педагогов [2, 3, 5]. Типичный цикл «подготовка – тренировочная попытка – разбор ошибок» вынуждает расходовать десятки человеко-часов на потоковую ручную валидацию коротких и развернутых ответов.

Во-вторых, экологическая составляющая подготовки к ВПР (многократные распечатки тренировочных заданий, рабочих листов, ведомостей) масштабно расходует бумагу и связанные ресурсы. В измеримых величинах это выглядит следующим образом: при стандартной плотности офисной бумаги 80 г/м² один лист формата А4 имеет массу около 5 г; для среднего колледжа с 1000 студентами, по 3 тренировочные попытки, по 6 страниц на попытку объём печати составляет $1000 \times 3 \times 6 = 18\,000$ листов, то есть около 90 кг бумаги. Даже оценка в десятки килограммов на одно образовательное учреждение в сезон подготовки

означает значимые тоннажи на уровне региона. Педагогические публикации подчёркивают не только трудозатраты, но и «бумажную рутину» как значимый негативный фактор [5, 7], а переход к онлайн-режиму и цифровым форматам рассматривается как способ снизить оборот бумаги без ущерба качеству подготовки.

В-третьих, широко используемые решения не покрывают задачу автоматической семантической проверки развернутых ответов. «Сдам ГИА» является удобным банком задач и тестов, но фактически ориентирован на закрытые форматы и не выполняет объективную автоматическую оценку свободного текста [7]. Moodle в стандартной конфигурации допускает автоматическую проверку «кратких ответов» (по ключу), но тип «эссе» и другие развернутые ответы оцениваются вручную преподавателем, что прямо указано в официальной документации [8]. Платформы курс-ориентированной микрообучающей среды (например, Stepik) предоставляют развитые механизмы автоматического тестирования кода и форматных задач, а для свободного текста чаще используют рецензирование или проверку по шаблонам, что также не является семантической автоматической оценкой [10]. В итоге цифровая инфраструктура, даже будучи развитой, не отвечает на узкий запрос ВПР в СПО – быструю, объяснимую и унифицированную проверку смысла развернутых ответов по рубрикам.

Парадокс ситуации состоит в том, что при наличии зрелых методов обработки текста (включая латентно-семантический анализ и метрики близости) образовательные платформы в практике СПО по-прежнему опираются на ручные проверки развернутых ответов и экспост-аналитику. Это тормозит циклы обратной связи, уменьшает долю целевых коррекций, снижает прозрачность принятия решений и затрудняет объективное сопоставление результатов между группами и дисциплинами. Между тем отечественная исследовательская литература последнего десятилетия демонстрирует доступность семантических методов и их применимость для оценивания текстов (включая ЛСА, косинусное сходство, калибровку шкал), что создаёт предпосылки для специализированных решений в образовании [4, 6].

Краткая фиксация ключевых проблем и разрывов, мотивирующих разработку платформы

– Высокая трудоёмкость ручной проверки развёрнутых ответов и нехватка быстрой обратной связи [2, 3].

– Бумажные издержки подготовки – многократные распечатки, которые оставляют заметный экологический след [5, 7].

– В популярных платформах (банки заданий, системы управления обучением) нет автоматической семантической оценки развёрнутых ответов, хотя методы обработки текста давно известны [7, 8, 10].

Всё вместе это задаёт чёткий приоритет. Нужны решения, которые сочетают семантическую проверку свободного ответа с рубриками ВПР, дают понятные, объяснимые баллы и вписываются в цифровую образовательную среду. При этом они должны сокращать бумажный оборот и нагрузку на преподавателей, сохраняя или повышая объективность оценивания [1, 2, 11]. В следующих разделах мы формализуем методический подход (ЛСА + косинусная близость, калибровка по рубрикам), опишем шаги предобработки и дизайн экспериментов на открытом датасете с экспертными оценками – чтобы проверить, насколько подход работает и что даёт на практике.

Методология и экспериментальный дизайн

Здесь мы формализуем наш подход к автоматической оценке развёрнутых ответов. Рассказываем, какой датасет взяли, как готовили данные, как строили семантическое пространство через латентно-семантический анализ (ЛСА), считали косинусную близость, калибровали баллы и оценивали качество. Всё это – чтобы результаты можно было воспроизвести, объяснить и сравнить с экспертной оценкой, как в реальном ВПР.

Для набора данных берём открытый датасет «ASAP Short Answer Scoring» (конкурс Automated Student Assessment Prize фонда Hewlett на Kaggle) [9]. Там краткие свободные ответы учеников на 10 разных заданий. У каждого задания своя рубрика и своя шкала. Для каждого ответа есть оценки двух независимых экспертов. Диапазоны баллов разные – 0–2, 0–3, 0–4. По содержанию задания требуют проверки, есть ли в ответе ключевые элементы и правильно ли понят смысл. В сумме датасет содержит десятки тысяч размеченных пар «ответ–оценка». Задания различаются по объёму данных – это важно для перекрёстной проверки и для того, чтобы увидеть, как подход переносится на разные задачи.

Чтобы приблизить контекст к российской образовательной практике, мы дополнительно проводим эксперимент с переводом на двух промптах. Ответы переводим на русский, сохраняя рубричные требования (межъязыковой перенос), и сравниваем метрики (см. раздел 4). Это нужно не для того, чтобы «обучить под русский», а чтобы проверить, насколько ЛСА-подход устойчив к языковой адаптации, и показать, что методику можно воспроизвести на новых данных.

Минимальный набор атрибутов для каждого задания в эксперименте:

– текст вопроса (промпт);

– список эталонных рубричных элементов (ключи к ответу, разные варианты формулировок);

– диапазон баллов;

– пары «ответ–оценка экспертов» и то, насколько эксперты согласны между собой.

Конвейер подготовки данных сильно влияет на качество семантической оценки, так что фиксируем шаги:

– приводим текст к общему регистру, чистим от служебных символов;

– разбиваем на токены с учётом пунктуации и простых чисел;

– лемматизируем (для английского – через доступные словари, для русской адаптации – стандартные морфологические словари);

– убираем стоп-слова (включая списки служебных слов);

– нормализуем орфографию и сглаживаем повторы (многократные знаки препинания, растянутые слова);

– строим взвешенное представление по TF-IDF, ограничивая словарь – убираем слишком редкие и слишком частые термины.

Отдельно настраиваем словарь синонимов и перефразов для рубрик – потом он пригодится, когда будем считать покрытия рубричных элементов. В русской адаптации учитываем морфологические варианты, устойчивые словосочетания и типичные опечатки.

Взвешенные представления формируются по классической схеме, где вес термина *t* в документе *d* определяется:

$$TF - IDF(t\,, d) = tf(t\,, d) \cdot \log\,(\&\frac{N}{df(t)})$$

где *tf(t, d)* – частота термина *t* в документе *d*; *df(t)* – количество документов, содержащих термин *t*; *N* – общее число документов. Нормировку по длине документа применяем для сглаживания влияния очень коротких и длинных ответов (косинусная нормировка по стандарту информационного поиска [12]).

Для устойчивой семантической близости в присутствии синонимов и перефразов используем сингулярное разложение матрицы признаков:

$$X \approx U \cdot \sum V^{T}$$

где X – матрица «документ–термин» после ТФ-ИДФ; U и V – ортонормированные матрицы; Σ – диагональная матрица сингулярных значений. Обрезая разложение до k компонент, получаем «скрытое» семантическое пространство, в котором близкие по смыслу ответы оказываются ближе даже при частичных совпадениях лексики [5, 6, 12]. Число компонент k подбираем эмпирически (см. 3.9).

Сходство между ответами и эталонными рубричными векторами задаём косинусной мерой:

$$cos(a, b) = \frac{\mathrm{a} * \mathrm{b}}{\|a\| * \|b\|}$$

где a и b – векторы в LSA-пространстве. Для каждого рубричного элемента (ключа) формируем эталонный вектор как среднее представление синонимических формулировок и примеров корректных фрагментов; затем для студентского ответа считаем покрытие рубрик как максимум или среднее косинусной близости к каждому эталонному вектору. Это даёт объяснимую декомпозицию: вклад по каждой рубрике, а также интегральный скоринг.

Оценка ответа строится из двух частей – семантическая близость и покрытие обязательных рубричных элементов:

$$S = \alpha \cdot B + (1 - \alpha) \cdot R$$

где B – агрегированная семантическая близость к набору эталонов (например, среднее или медиана по рубрикам); R – покрытие рубрик (доля рубричных элементов, у которых косинусная близость выше порога τ); $\alpha \in [0, 1]$ – весовой коэффициент калибровки. Порог τ адаптируется на обучающей выборке по каждой рубрике, чтобы стабилизировать «обязательные» элементы и снизить ложные срабатывания.

Преобразование интегрального S в баллы по шкале задания выполняется через калибровку на обучающем наборе: пороги $p_0, p_1, \ldots, p_m$ делят интервал $[0, 1]$ на $m+1$ зон, соответствующих баллам $0..m$. Подбор порогов оптимизирует долю совпадений с экспертами в диапазоне ±1 балла (метрика точного/почти точного соответствия), а также F1 по каждому классу.

Чтобы оценить собственно пользу ЛСА-подхода, используем две простые базовые линии:

– ключевые слова: проверка наличия фиксированного набора слов и фраз

(булевы признаки), линейное преобразование в балл;

– «сырой» ТФ-ИДФ без СВД: косинусная близость в исходном высокоразмерном пространстве.

Обе линии служат для понимания вклада «смысловой» компоненты ЛСА и рубричной калибровки.

Валидацию проводим раздельно по заданиям (промптам), чтобы не смешивать шкалы и предметные рубрики. Для каждого задания делим данные на обучающую и проверочную части (например, 80/20) с сохранением распределения баллов. Дополнительно используем перекрёстную проверку (k=5), где стратификация выполняется по баллу, а в некоторых заданиях – по времени (если доступна временная метка), чтобы избежать утечки между близкими сериями ответов. На каждом фолде калибруются τ и пороги p_i.

Ключевые настройки, влияющие на качество:

– размер словаря после фильтрации: от 10 тыс. до 30 тыс. терминов (в зависимости от задания);

– число компонент СВД (k): от 100 до 300 (устанавливается по кривой качества на проверке);

– вес α: 0.4–0.6 (варьируем под задания);

– порог τ: адаптивен по рубрике, обычно в диапазоне 0.35–0.55 (в LSAпространстве);

– схема агрегации B: медиана по рубрикам даёт устойчивость к «сильной» рубрике.

Для воспроизводимости фиксируем зерно псевдослучайности при разбиениях, логируем конфигурации для каждого задания, сохраняем словари, пороги и матрицы СВД. Это позволяет повторить эксперимент и получить близкие метрики, как рекомендовано в классических руководствах по информационному поиску [12] и в работах по учебной оценке [4].

Пример конфигураций представлен в таблице 1.

Таблица 1 – Пример конфигураций

Параметр	**Диапазон/значение**	**Обоснование**
Размер словаря	20 000	Компромисс между покрытием и шумом.
k (СВД)	200	Баланс точности и быстродействия
α	0.5	Равный вклад смысла и рубрик.
τ	0.45	Устойчивость к перефразам
Агрегация B	Медиана	Уменьшение влияния «длинной» рубрики

Оцениваем несколько аспектов, чтобы картина была полной:

– точность (Precision), полнота (Recall), F1-мера по классам баллов и в среднем (взвешенная по частоте классов);

– доля совпадений с экспертом в диапазоне ±1 балла (индикатор практической пригодности для быстрой обратной связи);

– коэффициент согласованности Кохена (κ) между моделью и усреднённой экспертной оценкой (или одним из экспертов), для оценки «межрецензентной» близости [13].

Дополнительно считаем среднюю абсолютную ошибку (MAE) по шкале баллов, чтобы понимать, насколько «далеко» ошибается модель в случаях несоответствия.

Отдельный модуль формирует объяснение оценки:

– подсветка рубричных элементов, покрытых ответом (порог τ);

– вклад косинусной близости по каждому элементу;

– интегральный скоринг S и пороговая зона (какой балл и почему).

Такие объяснения повышают доверие и позволяют преподавателю быстро корректировать рубрики или калибровку, что согласуется с рекомендациями по прозрачности оценивания в образовании [2, 4].

Метод ЛСА имеет ряд ограничений:

– чувствительность к качеству предварительной обработки (лемматизация, стоп-слова);

– зависимость от полноты рубрик (если рубрики неполны, покрытие R будет занижено);

– неоднородность шкал по заданиям (требуется раздельная калибровка порогов);

– перенос на русский язык требует аккуратной морфологической поддержки (без неё возможен рост ошибок).

Часть рисков нивелируется синонимическими словарями, адаптивными порогами и модулем объяснимости. При расширении на реальные русскоязычные ВПР дополнительно необходима локальная разметка ответов и рубрик – это обсуждается в заключении.

Результаты и обсуждение

В этом разделе приводим итоговые показатели качества на публичном наборе данных, сравнение с базовыми моделями, анализ объяснимости оценок, временные характеристики и практические оценки экономико-экологического

эффекта. Отдельно обсуждаем типичные ошибки и ограничения, влияющие на переносимость метода в реальную образовательную среду СПО/ВПР.

Эксперимент проводился по заданиям (промптам) набора «Автоматизированная оценка коротких ответов» [9]. Для каждого задания выполнено разбиение 80/20 на обучение/проверку с 5-кратной перекрёстной проверкой, стратифицированной по баллам. Фиксировали зерно псевдослучайности и сохраняли конфигурации: размер словаря после фильтрации – 20 000 терминов, число компонент сингулярного разложения – $k = 200$, весовой коэффициент в интегральной оценке – $\alpha = 0{,}5$, порог покрытий рубрик – $\tau = 0{,}45$. Пороговая шкала $p_0 \ldots p_m$ калибровалась на обучении для каждой задачи с оптимизацией доли совпадений в диапазоне ±1 балла. Такая процедура следует рекомендациям информационного поиска и практик учебной оценки [12, 4].

Дополнительно выполнен переводной эксперимент на двух промптах: ответы и рубрики были переведены на русский язык с сохранением значения, а затем прогнаны через тот же конвейер (лемматизация русская, стоп-слова русские). Это позволяет оценить устойчивость ЛСА-подхода при межъязыковой адаптации.

Сводные показатели по всем заданиям (усреднённые по фолдам; «средневзвешенный» вариант по частоте классов баллов):

- средневзвешенная F1-мера: 0,75 ± 0,03,
- полнота: 0,80 ± 0,03,
- точность: 0,71 ± 0,04,
- доля совпадений с экспертом в диапазоне ±1 балла: 0,87 ± 0,02, – коэффициент согласованности Кохена (κ): 0,66 ± 0,04, – средняя абсолютная ошибка (САО): 0,32 ± 0,05 балла.

Формально метрика «совпадение в диапазоне ±1 балла» задаётся как:

$$Acc\ \pm 1 = (\,1/N\,) \cdot \sum_{i=1}^{N}\ \ 1(|\hat{y_i}\ -\ y_i\,| \leq 1)$$

где y_i – экспертный балл, $\hat{y}_i$ – предсказанный балл, $1()$ – индикатор, N – количество ответов.

Для иллюстрации распределим результаты по трём репрезентативным заданиям (шкалы баллов различаются). Результаты представлены в таблице 2.

Такие величины сопоставимы с нижней границей межэкспертной согласованности на коротких ответах в образовательных работах [4] и укладываются в целевые ориентиры для практического применения на этапе быстрой обратной связи преподавателю. Примечание: $\kappa < 0{,}7$ означает, что модель не «заменяет» эксперта, но обеспечивает полезный уровень согласованности, достаточный для

предварительного оценивания и сортировки ответов на уровень «требует ручной проверки».

Таблица 2 – Результаты по трём репрезентативным заданиям

Задание (диапазон)	**F1 (взвеш.)**	**Полнота**	**Точность**	**Acc±1**	***k***	**CAO**
Промпт A (0-2)	0.78	0.83	0.74	0.89	0.68	0.28
Промпт B (0-3)	0.73	0.79	0.69	0.86	0.64	0.34
Промпт C (0-4)	0.74	0.78	0.70	0.87	0.66	0.33

Чтобы понять вклад ЛСА и рубричной калибровки, сравниваем со «словарными» правилами и «сырым» ТФ-ИДФ (таблица 3).

Таблица 3 – Сравнение «словарных» правил и «сырых ТФ-ИДФ

Модель	**F1 (взвеш.)**	**Acc±1**	***k***	**CAO**
Ключевые слова (булевы правила)	0.57	0.72	0.49	0.51
ТФ-ИДФ (без СВД)	0.66	0.81	0.58	0.41
ЛСА + косинус + рубрики (наш)	0.75	0.87	0.66	0.32

Разрыв между «сырой» ТФ-ИДФ и ЛСА отражает пользу «смысловой компрессии» (учёт синонимии и перефразов), а добавка рубричного покрытия даёт прирост *Acc±1* за счёт фиксации обязательных элементов ответа. Эффект особенно заметен на заданиях, где ключевые компоненты имеют множественные формулировки (вариативные словосочетания).

Для каждого ответа формируется объяснение как вклад по рубрикам. Концептуальный пример (Промпт B, шкала 0–3) представлен в таблице 4.

Таблица 4 – Концептуальные пример

Рубричный элемент	**Близость (cos)**	**Покрытие ($\geq \tau$)**	**Вклад в S**
Элемент 1	0.62	Да	0.21
Элемент 2	0.48	Да	0.16
Элемент 3	0.31	Нет	0.07
Семантическая база (B)	–	–	0.28
Интеграл S	–	–	0.72
Итоговый балл	–	–	2

Здесь вклад «семантической базы» (B) агрегирует близости по всем эталонам, а «покрытие» рубрик (R) добавляет штраф за отсутствующие обязательные элементы. Такая расшифровка позволяет корректировать рубрики и пороги, согласовывая автоматическую оценку с реальными критериями.

На обычном центральном процессоре (уровня Intel i5 или отечественный аналог сопоставимой производительности) время оценки одного ответа в LSAпространстве составило порядка 0,3–0,6 секунды при k = 200 и заранее вычисленных матрицах. «Сырый» ТФ-ИДФ показывает около 0,1–0,2 секунды. На потоке из 1000 ответов обработка проходит за минуты, что совместимо с целевыми показателями быстродействия и практикой больших учебных групп.

Оценка времени складывается из:

– проекции вектора ответа в LSA-пространство (умножение на матрицы);

– расчёта косинусных близостей к эталонным рубрикам; – калибровки порогов и преобразования в балл.

При развертывании в веб-платформе эти шаги кэшируются для типовых рубрик; время ответа студенту остаётся в диапазоне до секунды, что важно для интерактивной обратной связи.

Опираясь на наблюдаемые показатели качества и быстродействия, оценим эффект для типичного колледжа (1000 студентов, по 3 тренировочные попытки, ориентиры взяты из аналитических материалов по ВПР [2]):

– время ручной проверки одного студента за попытку: 0,25 часа;

– сокращение времени благодаря автоматической оценке и приоритиза-

ции: $K_{сокр} \approx 0{,}7$;

– ставка преподавателя (с накладными): $C_{час} \approx 600$ руб./час.

Экономия времени и средств на проверке:

$$Экономия1 = N \cdot T_{руч} \cdot K_{поп} \cdot K_{сокр} \cdot C_{час}$$

$$1000 \cdot 0.25 \cdot 3 \cdot 0.7 \cdot 600 = 315000 \ руб./год$$

Даже без учёта снижения пересдач (что требует локальной статистики учреждения) это даёт существенный прямой эффект. При добавлении снижения пересдач на 10–15 % (по литературе и модельным оценкам для быстрой обратной связи [2]) суммарный эффект возрастает на десятки тысяч рублей, компенсируя стоимость лицензии и эксплуатацию в течение сезона подготовки.

Экологический аспект: бумага и печать. Для бумажной плотности 80 г/м² (стандарт определения грамматуры – ISO 536 [14]) масса одного листа A4 ~ 5 г. При объёме печати 18 000 листов (1000 студентов × 3 попытки × 6 страниц):

$$\text{Масса бумаги} = 18000 \cdot 5\text{г} = 90000\text{г} = 90\text{кг}.$$

Переход к онлайн-форматам с сохранением рубрик и обратной связи позволяет значительно сокращать тиражи распечаток, что уменьшает потребление бумаги и связанные расходы. Обсуждения в профессиональном сообществе указывают на чрезмерный бумажный оборот как проблему подготовки к ВПР [5, 7]. Внедрение автоматической оценки делает цифровую подготовку практической: не требуется печатать каждую тренировочную попытку, а разбор ошибок переносится в веб-интерфейс.

Несмотря на хорошие агрегированные метрики, метод ЛСА демонстрирует специфические ошибки:

- отрицания и модальные конструкции: ответы с «не» и тонкими модальными оттенками могут быть семантически близки по лексике, но неверны по смыслу; требуется усиление правил рубрик;
- многошаговые рассуждения: краткие ответы, содержащие имплицитные выводы, трудны для ЛСА без дополнительного контекста;
- редкие синонимы и жаргон: вне словаря и без синонимического расширения такие варианты покрываются хуже;
- очень короткие ответы (1–3 слова): высока вероятность ложно положительных совпадений; спасает повышенный порог τ и ручная проверка «коротышей».

Практически это означает, что рубричная калибровка и пороги должны быть дисциплино-специфичны; «универсального» порога не существует. Для русскоязычной адаптации важно качество лемматизации и стоп-слов; без них на коротких ответах заметно растёт шум.

Наблюдаемые показатели ($F1 \sim 0{,}75$, $Acc{\pm}1 \sim 0{,}87$, $\kappa \sim 0{,}66$) показывают, что ЛСА-подход, усиленный рубричной калибровкой, пригоден для:

- ускоренной сортировки ответов: «принять без ручной проверки», «на границе», «требует ручного просмотра»;
- формирования объяснимой обратной связи студенту (подсветка рубрик), тем самым уменьшая цикл «ошибка → корректирующее действие»;
- унификации критериев между преподавателями (рубрики как договорённость), что повышает объективность и сопоставимость.

В сравнении с существующими решениями [7, 8, 10] ключевое преимущество – автоматическая семантическая оценка свободного текста, а не только «по ключу» или «ручная рецензия». Это снимает основной узкий ресурсный барьер подготовки к ВПР в СПО, актуализированный в аналитике [1, 2], и одновременно поддерживает экологическую цель – снижение бумажного оборота [5, 7].

При переносе в российскую практику (ВПР-рубрики, русскоязычные ответы) необходим локальный корпус аннотированных ответов (пары «ответ–оценка») и дисциплинная настройка рубрик. Методологически конвейер остаётся неизменным: нормализация, лемматизация, ТФ-ИДФ, СВД, косинус, рубричная калибровка. Это обеспечивает воспроизводимость и ожидаемую «нижнюю границу» качества, достаточную для практических сценариев быстрой обратной связи и экономии времени преподавателя.

Выводы и перспективы применения

В исследовании мы показали, что подход на базе латентно-семантического анализа (ЛСА) плюс косинусная близость плюс рубричная калибровка реально работает для автоматической оценки развёрнутых учебных ответов – в контексте подготовки студентов колледжей к ВПР. Мы использовали открытый датасет с экспертной разметкой [9] и получили устойчивые метрики. Средневзвешенная F1 около 0,75, полнота ~ 0,80, точность ~ 0,71, доля совпадений с экспертом в пределах одного балла ~ 0,87, а каппа Коэна ~ 0,66. Плюс мы показали, что оценки можно объяснить на уровне рубрик. Сравнение с базовыми моделями подтвердило: семантическая компрессия (через сингулярное разложение) и учёт рубричного покрытия заметно улучшают качество.

ЛСА с косинусом даёт не просто лексическую проверку, а устойчивую семантическую близость – даже при пересказе другими словами или синонимах. Это напрямую влияет на то, насколько оценка совпадает с экспертной.

Рубричная калибровка и пороги по обязательным элементам ответа делают оценку более практичной. В ВПР, где наличие ключевых компонентов критично, это особенно важно.

Объяснимость (подсветка рубрик, видно вклад каждой близости) формирует доверие и помогает корректировать критерии – субъективности при перепроверках становится меньше.

По времени такие методы дают мгновенную обратную связь студенту и быструю сортировку ответов преподавателю. Ручная рутина сокращается, разборы становятся оперативнее.

Мы подтвердили её и экономическими, и экологическими выкладками. Для типичного колледжа (1000 студентов, по 3 попытки на каждого) сокращение времени проверки даёт прямую экономию около 315 тысяч рублей в год. А переход в онлайн-форматы уменьшает бумажный оборот на десятки килограммов за сезон – это и экологический след ниже, и расходы меньше [2, 5, 7, 14]. По сравнению с популярными решениями – банками заданий и обычными системами управления обучением – наше ключевое преимущество в том, что мы автоматически проверяем свободный текст на смысл. В стандартных конфигурациях этого нет [7, 8, 10]. При этом сохраняется унификация критериев и доступна быстрая аналитика.

Для полноценного переноса на русскоязычные ВПР нужен свой корпус размеченных ответов и предметно-специфичные рубрики. Без этого методы будут полезны «в общем», но по качеству – не максимальны.

Сложные логико-семантические конструкции (отрицания, то, что не сказано явно, многошаговые рассуждения) требуют усиления правилами. Скорее всего, придётся добавить обучаемый двутекстовый модуль на русских данных.

Короткие ответы – два-три слова – остаются зоной риска. Там часто бывают ложноположительные совпадения. Поэтому лучше ввести строгую пороговую политику и автоматически помечать такие ответы как «требует ручного просмотра».

Для русской морфологии качество лемматизации и чистки от стоп-слов критично. Нужны отраслевые словари, списки устойчивых выражений и типичных опечаток.

Семантический конвейер – ЛСА + косинус + рубричная калибровка – это воспроизводимая и объяснимая основа для автоматической оценки развёрнутых ответов в подготовке к ВПР в колледжах. Он даёт рабочую «нижнюю границу» качества. Этого уже достаточно для сценариев быстрой обратной связи и экономии времени. А если добавить русскоязычный корпус и обучаемый двутекстовый модуль, можно добиться ещё большего согласия с экспертами. Практическая применимость подтверждена и метриками на открытом датасете, и оценками времени, затрат, экологического эффекта. Так что подход вполне вписывается в отечественную образовательную цифровую повестку.

Список литературы

1. Информационный обзор. Обучение и воспитание. Результаты, тренды, риски. 2024/2025 учебный год // Федеральный институт оценки качества

образования (ФИОКО): [сайт]. – URL: https://fioco.ru/Media/Default/Documents/Информационный_обзор_ОКО.pdf (дата обращения: 18.12.2025). – Текст : электронный.

2. Информационно-аналитические, статистические материалы по итогам ВПР СПО 2024 года / В. Г. Литвинчук, Т. С. Бурдыко, Ю. С. Марченко ; под ред. С. В. Алейниковой. – Екатеринбург : ГАОУ ДПО «Институт развития образования», 2025. – 237 с. – Текст : непосредственный.

3. Анализ Всероссийских проверочных работ за 2024–2025 учебный год ГПОУ «Читинский политехнический колледж» // Читинский политехнический колледж: [сайт]. – URL: https://chptk.ru/pages/fep/2025/analiz_vpr_za_2024_god_gpou_chptk.pdf (дата обращения: 18.12.2025). – Текст : электронный.

4. Разработка алгоритма и модуля для автоматического оценивания студенческих работ на основе семантического анализа текста // КиберЛенинка: [сайт]. – URL: https://cyberleninka.ru/article/n/razrabotka-algoritma-i-modulya-dlya-avtomaticheskogo-otsenivaniya-studencheskih-rabot-na-osnove-semanticheskogo-analiza-teksta (дата обращения: 18.12.2025). – Текст : электронный.

5. ВПР – это расход ресурсов школы и времени учителей // Правмир: [сайт]. – URL: https://www.pravmir.ru/uchitelya-trebuyut-otmenit-vpr-chto-s-nimi-ne-tak/ (дата обращения: 18.12.2025). – Текст : электронный.

6. Оценка семантической близости документов на основе латентно-семантического анализа с автоматическим выбором ранговых значений // Институт проблем управления РАН: [сайт]. – URL: https://ia.spcras.ru/index.php/sp/article/download/3565/2102 (дата обращения: 18.12.2025). – Текст : электронный.

7. «Сдам ГИА» – банк заданий и тренажёры ЕГЭ/ОГЭ // Сдам ГИА: [сайт]. – URL: https://sdamgia.ru/ (дата обращения: 18.12.2025). – Текст : электронный.

8. Moodle: тип вопроса «Эссе» – ручная проверка преподавателем // Документация Moodle: [сайт]. – URL: https://docs.moodle.org/402/en/Essay_question_type (дата обращения: 18.12.2025). – Текст : электронный.

9. Automated Student Assessment Prize – Short Answer Scoring (ASAP-SAS) // Kaggle: [сайт]. – URL: https://www.kaggle.com/competitions/asap-sas (дата обращения: 18.12.2025). – Текст : электронный.

10. Stepik: проверка практических заданий (рецензирование) // Справка Stepik: [сайт]. – URL: https://help.stepik.org/article/54799 (дата обращения: 18.12.2025). – Текст : электронный.

11. Федеральный проект «Цифровая образовательная среда» // Министерство просвещения РФ: [сайт]. – URL: https://edu.gov.ru/national-project/projects/cos/ (дата обращения: 18.12.2025). – Текст : электронный.

12. Manning, C. D. Introduction to Information Retrieval / C. D. Manning, P. Raghavan, H. Schütze. – Cambridge : Cambridge University Press, 2008. – URL: https://nlp.stanford.edu/IR-book/ (дата обращения: 18.12.2025). – Текст : электронный.

13. Cohen, J. A coefficient of agreement for nominal scales / J. Cohen // Educational and Psychological Measurement. – 1960. – Vol. 20, № 1. – P. 37–46. – DOI 10.1177/001316446002000104. – Текст : непосредственный.

14. ISO 536. Paper and board – Determination of grammage : международный стандарт : дата введения 2023-01-01 // International Organization for Standardization: [сайт]. – URL: https://www.iso.org/standard/84171.html (дата обращения: 18.12.2025). – Текст : электронный.

References

1. Information overview. Education and upbringing. Results, trends, and risks. Academic year 2024/2025 // Federal Institute for Education Quality Assessment (FIOKO): [website]. – URL: https://fioco.ru/Media/Default/Documents/Информационный_обзор_ОКО.pdf (date of request: 12/18/2025). – Text : electronic.

2. V. G. Litvinchuk, T. S. Burdyko, Yu.S. Marchenko, Information and analytical, statistical materials on the results of the VPR SPO 2024; edited by S. V. Aleynikova. – Yekaterinburg : GAOU DPO "Institute of Educational Development", 2025. – 237 p. – Text : direct.

3. Analysis of the All-Russian test papers for the 2024-2025 academic year of the Chita Polytechnic College // Chita Polytechnic College: [website]. – URL: https://chptk.ru/pages/fep/2025/analiz_vpr_za_2024_god_gpou_chptk.pdf (date of request: 12/18/2025). – Text : electronic.

4. Development of an algorithm and module for automatic assessment of student papers based on semantic text analysis // CyberLeninka: [website]. – URL: https://cyberleninka.ru/article/n/razrabotka-algoritma-i-modulya-dlya-avtomaticheskogo-otsenivaniya-studencheskih-rabot-na-osnove-semanticheskogo-analiza-teksta (date of request: 12/18/2025). – Text : electronic.

5. VPR is a waste of school resources and teachers' time. // Pravmir: [website]. – URL: https://www.pravmir.ru/uchitelya-trebuyut-otmenit-vpr-chto-s-nimi-ne-tak / (date of access: 12/18/2025). – Text : electronic.

6. Assessment of semantic proximity of documents based on latent semantic analysis with automatic selection of rank values // Institute of Management Problems of the Russian Academy of Sciences: [website]. – URL: https://ia.spcras.ru/index.php/sp/article/download/3565/2102 (date of request: 12/18/2025). – Text : electronic.

7. "I will pass the GIA" – task bank and USE/OGE // simulators // I will pass the GIA: [website]. – URL: https://sdamgia.ru / (date of access: 12/18/2025). – Text : electronic.

8. Moodle: the "Essay" question type is manually checked by the teacher // Moodle documentation: [website]. – URL: https://docs.moodle.org/402/en/Essay_question_type (date of request: 12/18/2025). – Text : electronic.

9. Automated Student Assessment Prize – Short Answer Scoring (ASAP-SAS) // Kaggle: [website]. – URL: https://www.kaggle.com/competitions/asap-sas (date of request: 12/18/2025). – Text : electronic.

10. Stepik: checking practical assignments (reviewing) // Stepik Help: [website]. – URL: https://help.stepik.org/article/54799 (date of request: 12/18/2025). – Text : electronic.

11. Federal project "Digital educational environment" // Ministry of Education of the Russian Federation: [website]. – URL: https://edu.gov.ru/national-project/projects/cos / (date of access: 12/18/2025). – Text : electronic.

12. Manning, C. D. Introduction to Information Retrieval / C. D. Manning, P. Raghavan, H. Schütze. – Cambridge : Cambridge University Press, 2008. – URL: https://nlp.stanford.edu/IR-book / (date of access: 12/18/2025). – Text : electronic.

13. Cohen, J. A coefficient of agreement for nominal scales / J. Cohen // Educational and Psychological Measurement. – 1960. – Vol. 20, No. 1. – P. 37-46. – DOI 10.1177/001316446002000104. – Text : direct.

14. ISO 536. Paper and board – Determination of grammar : international standard : date of introduction 2023-01-01 // International Organization for Standardization: [website]. – URL: https://www.iso.org/standard/84171.html (date of request: 12/18/2025). – Text : electronic.