

ОРИГИНАЛЬНАЯ СТАТЬЯ

УДК 378.147: 519.767.6

doi: 10.26907/2541-7746.2024.4.580-593

**АВТОМАТИЗИРОВАННАЯ ОЦЕНКА ОТВЕТОВ
ПРИ КОНТРОЛЕ ЗНАНИЙ ДЛЯ ВОПРОСОВ ТИПА
«ОПРЕДЕЛЕНИЕ» И «ОПИСАНИЕ»**

Н.А. Прокопьев

Казанский (Приволжский) федеральный университет, г. Казань, 420008, Россия

Аннотация

Разработка эффективных систем автоматизированного контроля знаний непосредственно связана с реализацией программного модуля для оценки ответов на вопросы контрольного теста, свободно сформулированных на естественном языке. Ранее автором был разработан экспериментальный прототип такой системы и проведен эксперимент, в котором короткие ответы на вопросы базовых типов, полученные от студентов, были обработаны прагматически-ориентированным алгоритмом обработки вопросно-ответных текстов для оценки результатов работы алгоритма, нахождения и разрешения ситуаций некорректного функционирования. В настоящей статье представлена следующая итерация развития алгоритма для обработки ответов на более сложные вопросы, подразумевающие раскрытие нескольких отношений, такие как «Определение» и «Описание». Представлено описание особенностей алгоритма, в частности, решение проблемы сегментации текста ответа на базовые части, задача обработки которых уже решена. Описаны результаты эксперимента, проведенного с использованием полученного прототипа с получением ответов от студентов и анализом ситуаций некорректного функционирования.

Ключевые слова: обработка естественного языка, автоматизированная оценка ответов, цифровое образование

Введение

В настоящее время наблюдается усиливающаяся тенденция масштабного перехода к дистанционному образованию и удаленному управлению образовательным процессом, что смещает акценты функциональной ответственности в трехагентной образовательной модели «Ученик – Учитель – Технология» [1]. В результате третий агент в этой модели, изначально игравший роль вспомогательного ресурса, получает более высокий статус, частично перенимая функции второго агента – Учителя. Следовательно, существует необходимость разработки и внедрения в системы цифрового образования программного модуля для непрерывного интеллектуального анализа качества обучения Ученика и получения эффективной обратной связи. Решение задачи разработки интеллектуальных программных модулей для анализа и оценки ответов Ученика, данных им в свободной форме и на естественном языке, становится наиболее актуальным, поскольку может значительно улучшить качество оценок уровня знаний и усвоения предмета.

Ранее нами был реализован алгоритм обработки естественно-языковых ответов для случаев, когда ответ имеет достаточно простую структуру, раскрывающую одно типовое или составное отношение (в терминах семантического анализа). Были разработаны основные методологические принципы реализации, создан экспериментальный прототип системы автоматизированной оценки знаний, реализующий предложенный алгоритм. Полученные результаты представлены в [2]. Кроме того, была реализована система контроля знаний, с помощью которой были получены ответы студентов на вопросы в рамках аутентичного контроля знаний. Названный выше алгоритм верифицирован на полученных ответах экспериментально, также выявлены ситуации некорректного функционирования этого алгоритма и пути их разрешения. В результате была подтверждена гипотеза о возможности построения эффективного прагматически-ориентированного алгоритма обработки вопросно-ответных текстов. Эти результаты представлены в статье [3].

Однако многие задаваемые вопросы, в частности, такие как «Что такое X?», «Дайте определение X» или «Опишите X», где X – некоторый концепт учебной дисциплины, требуют более сложного, составного ответа, включающего раскрытие некоторого набора отношений, анализ которых не был реализован в названном выше экспериментальном прототипе системы автоматизированной оценки знаний. В настоящей статье представлено дальнейшее развитие этого прототипа в указанном направлении. В частности, показано, что анализ ответов на вопросы названных типов может быть сведен к рекурсивному анализу отдельных сегментов ответа с использованием уже реализованных методов для типовых и составных отношений при условии решения задачи семантико-синтаксической сегментации текста.

1. Методология исследования

Основой методологических принципов подхода, использованного для построения прототипа системы автоматизированной оценки знаний, является тезис, что заданный вопрос естественным образом ограничивает контекст ответа как в отношении вариативности возможных ответов, так и его структуры. Прагматически-ориентированный подход, представленный в [4], основан на следующих принципах:

1. Естественная ограниченность пространства значений ответа и соответствующих лексем по заданному вопросу в учебно-тестовой ситуации (принцип детерминированности значений ответа).
2. Естественная ограниченность типов структур ответа заданным вопросом, описываемых в терминах специальных глубинных грамматик (принцип детерминированности формы ответа).
3. Возможность управления точностью оценки ответа в силу расширяемости и модифицируемости глубинных грамматик как основы прагматически-ориентированного алгоритма оценки (принцип открытости грамматик).
4. Возможность сужения пространства действия алгоритма оценки за счет сравнения ответа обучаемого с заранее заготовленной моделью ответа (принцип ожидаемости ответа).

Основными понятиями в теоретической модели являются «Концептула» и «Индивидуальная концептуальная грамматика». Концептула выражает конкретную типизированную семантико-грамматическую роль лексем или их частей в ответе, указывая на соответствующие грамматические признаки естественного языка.

Индивидуальная концептуальная грамматика (ИКГ) представляет собой схему сочетания концептул, соответствующую правильной передаче ожидаемого смысла ответа в соответствии с его семантическим классом. Каждая ИКГ передает некоторый канонический смысл, а ответ, соотносящийся с этим смыслом, называется канонизированным ответом. Таким образом, необходимость введения ИКГ заключается в сведении семантического анализа текста к синтаксическому анализу его канонизированного представления в условиях, определенных контекстом вопроса.

Алгоритм обработки ответов представлен схематически на рис. 1. Лексический процессор получает на вход ответ, выраженный на естественном языке, и модель ответа для заданного вопроса. Модель ответа представляет собой словарь соответствий вида «Концептула» – «Список лексем, соответствующих концептуле». Лексемы из модели ответа могут содержать символы * (обозначает произвольный одиночный символ) и & (обозначает произвольную последовательность символов). Эти символы позволяют автору вопроса предусмотреть различные словоформы и учесть синтаксические неточности в ответе обучаемого. Пример модели ответа приведен далее в разделе, посвященном реализации прототипа.

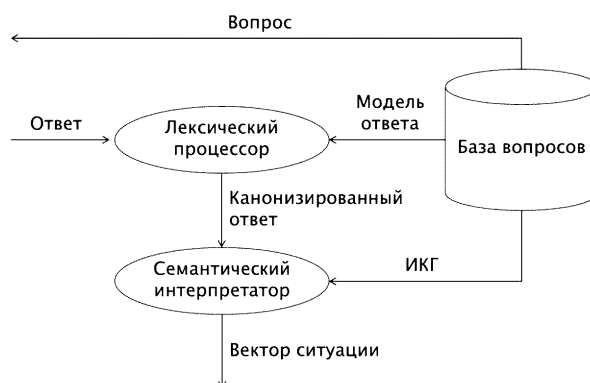


Рис. 1. Алгоритм обработки ответов

На выходе лексический процессор формирует канонизированное представление ответа – цепочки концептул, соответствующих естественно-языковой форме ответа. Для представления ответа в виде цепочки концептул лексический процессор производит разбиение ответа на лексем и выполняет поиск соответствующих концептул в модели ответа. Пример канонизированного представления приведен далее в разделе «Реализация и тестирование прототипа». Кроме того, лексический процессор формирует частичный вектор ситуации, массив запрещенных лексем и массив неопределенных лексем из ответа. Частичный вектор ситуации – это числовой вектор $(S1, S2, S3, S4)$:

- $S1$ – отношение количества лексем в ответе к количеству лексем, предусмотренных моделью ответа;
- $S2$ – количество запрещенных лексем в ответе;
- $S3$ – количество неопределенных лексем в ответе;
- $S4$ – код, характеризующий модальность ответа, т.е. либо наличие в ответе лексем, характеризующих неуверенность или категоричность, либо их отсутствие, т.е. нейтральность. Значением $S4$ являются 0, 1 или 2 для случаев неуверенности, категоричности или нейтральности соответственно.

На вход семантического интерпретатора поступают канонизированное представление ответа и ИКГ, соответствующая заданному вопросу. Интерпретатор проверяет соответствие ответа и ИКГ путем попытки обхода синтаксического графа ИКГ по узлам согласно цепочке концептуал канонизированного ответа. Если обход завершается на конечном узле графа, то считается, что ответ соответствует ИКГ. На выходе семантический интерпретатор формирует полный вектор ситуации ($S_1, S_2, S_3, S_4, S_5, S_6, S_7$), который в дополнение к частичному вектору ситуации содержит:

- S_5 – код, характеризующий правильность использования интервальных лексем; принимает значение 0, если интервальные лексемы использованы верно, иначе принимает значение 1;
- S_6 – код, характеризующий правильность смысла ответа, т. е. его соответствие ИКГ вопроса; его значением является 0, если канонизированный ответ соответствует ИКГ; если в ответе пропущено отношение, то принимает значение 1; иначе, если канонизированный ответ не соответствует ИКГ, принимает значение 2;
- S_7 – код, характеризующий полноту ответа, т. е. соотношение длины канонизированного представления и частично либо полностью соответствующей ей цепочки концептуал из ИКГ; принимает значение 0 при полном соответствии; если канонизированный ответ длиннее соответствующей цепочки, то принимает значение 1, иначе принимает значение 2.

Гипотеза о возможности построения эффективного прагматически-ориентированного алгоритма для обработки естественно-языковых вопросно-ответных текстов была подтверждена в [3] в рамках экспериментального исследования для вопросов, подразумевающих раскрытие одиночных типовых и составных отношений. В настоящем исследовании использован тот же подход, который состоит в проведении аутентичного тематического контроля знаний и анализе полученных ответов студентов с использованием разработанного прототипа.

Для обработки более сложных по структуре ответов на вопросы типа «Определение» и «Описание» предложен подход, состоящий в рекурсивном применении уже реализованных программных модулей для семантического анализа, поскольку ответы на такие вопросы подразумевают раскрытие некоторого набора типовых и составных отношений, связанных с главной сущностью вопроса. Основной проблемой в данном случае является сегментация текста, т. е. его разделение на такие фрагменты, которые полностью соответствуют конкретному отношению и к которым применимы соответствующие ИКГ. Такая сегментация текста также может быть произведена согласно прагматически-ориентированному подходу в соответствии с методом, предложенным в [4].

В естественном языке минимальной смысловой конструкцией является предложение, однако оно само может включать отдельные смысловые части, т. е. может являться составной конструкцией. Поэтому необходимы не только синтаксический, но и семантический подходы к сегментации. При реализации семантической сегментации возникают вопросы выделения контекста, в рамках которого текст ответа должен анализироваться на смысловую корректность, а также должны выделяться именно такие сегменты, к которым можно применить простые формулы ИКГ. Принципы ожидаемости ответа и детерминированности контекста (структуры и смысла ответа) позволяют разрешить эту проблему, так как можно заранее задать главное понятие вопроса, вокруг которого строится контекст ответа, а также определить рекурсивную ИКГ, к которой ответ должен относиться.

Новый сегмент определяется по следующим признакам:

1. Синтаксический, поверхностный признак. Обозначается в тексте явно: знаком (символы пунктуации) или конкретной функциональной лексемой (лексемы типа «который», «что», «такой, что» и т. п.).
2. Семантический, глубинный признак. Содержательно определяет сегмент и представляется лексемой, отражающей новое отношение как между главным понятием и другим понятием (линейная структура), так и взаимно между другими понятиями (иерархическая структура).

Процесс выделения сегментов одновременно с канонизацией текста лексическим процессором и анализом его соответствия простым формулам семантическим интерпретатором продолжается рекурсивно до завершения текста ответа.

2. Системы обработки естественно-языковых ответов

Для сравнения с существующими подходами дадим обзор систем, разработанных в области обработки естественно-языковых ответов.

В [5] дан анализ систем автоматизированной оценки коротких ответов на вопросы. Понятие «короткий ответ на вопрос» авторы определяют по следующим пяти критериям:

1. Ответ на заданный вопрос невозможно извлечь из самого вопроса, т. е. ответ может быть сформирован только из внешних знаний, не извлекаемых из вопроса.
2. Ответ должен выражаться на естественном языке.
3. Длина ответа должна быть не более одного абзаца (от одного предложения до одного абзаца).
4. Оценивается прежде всего смысл ответа и в меньшей степени его форма – критерий приоритета извлечения смысла.
5. Вариативность ответа должна ограничиваться за счет контекста вопроса.

Авторы названной статьи провели исторический анализ развития систем оценки коротких ответов за период с 1995 по 2014 год, предложив разделить этот временной промежуток на пять периодов, названных эпохами. Эпохи могут пересекаться, при этом каждая эпоха характеризуется появлением и развитием определенных идей и подходов к оценке коротких ответов и некоторым уровнем прикладных исследований и программных прототипов, реализованных в этом направлении. При сравнительном анализе систем оценки коротких ответов, разработанных за рассматриваемый период времени, авторы использовали шесть характеристик таких систем: набор данных, метод обработки естественного языка, модель данных, модель оценки ответов, оценка качества модели и оценка эффективности.

Наборы данных в большинстве систем формируются из академического и преподавательского опыта, открытых источников либо путем непосредственного проведения тестирования. Основные языки наборов данных – английский, китайский, немецкий и испанский. Часто встречающиеся области знаний для тестирования – информатика, естественные науки, языкознание.

Для обработки естественного языка в различной комбинации применяются 17 методик, наиболее частыми при этом являются: проверка орфографии, стеммирование и частеречная разметка. Наименее частыми являются синтаксические шаблоны, разметка именованных сущностей и сегментация текста.

На основе исследованных моделей, в частности, соответствия вопросов, ответов обучаемых и эталонных ответов, авторы сделали вывод, что эталонные ответы играют разную роль в системе оценки ответов: в одних случаях эталонный ответ и есть модель ответа, в других он используется для разметки ответа обучаемого.

Модели оценки ответов разнятся в зависимости от эпохи и используемых подходов и разбиваются на два больших класса: оценки на основе правил (эпоха карт концептов и извлечения информации) и оценки статистическими методами (эпоха использования корпусов и машинного обучения). Согласно названным исследованиям, наиболее широко распространенными являются системы с машинным обучением.

Определение качества моделей в рассмотренных системах зависит от типа модели оценки ответа. Оценка на основе правил лучше всего реализуется в повторяющемся тестировании по одной и той же области знаний, известной заранее. Оценка на основе статистических данных лучше всего показывает себя в неповторяющемся тестировании по области знаний, неизвестной заранее. Авторы также выделили три класса используемых метрик для оценки качества моделей: классифицирующие, ранжирующие и метрики отношений.

Основной вывод авторов о способах оценки эффективности рассмотренных систем заключается в следующем: оценка усложнена отсутствием общих наборов данных и единых способов оценки моделей. Поэтому пока лучше всего показывает себя статистический подход к оценке эффективности по принципам TREC Eval.

В [6] приведены результаты анализа систем проверки ответов обучаемого, а рассматриваемые системы разделены на три класса: системы на основе шаблонов-масок; системы на основе модели «мешок слов»; системы, учитывающие ролевые функции слов.

Системы на основе шаблонов-масок характеризуются тем, что эталонные ответы в них задаются на некотором логическом языке шаблонов и описывают варианты ответа на вопрос. В качестве основного недостатка таких систем авторы выделяют высокую нагрузку на преподавателя, приводящую, в том числе, к стремлению сократить шаблоны, из-за чего снижается точность оценки ответов.

Системы, основанные на модели «мешок слов», используют в основном методы статистического анализа и поиск n -грамм в ответе с предварительной обработкой ответа: исключение парафразы, снятие синонимии, расшифровка аббревиатур. К таким системам авторы относят системы, использующие векторные модели представления ответа и методы машинного обучения. Плюсом таких систем является относительная простота подготовки контрольных тестов, однако для случая вопросов, требующих определенную структуру ответа, например, требующих перечислить этапы некоторого процесса в нужном порядке, такие системы не подходят.

Системы, учитывающие ролевые функции слов в предложении и их взаимную связь, комбинируют подходы двух других названных классов систем, так как совмещают структурный подход и гибкость построения модели ответа. Однако на данный момент времени такие системы, согласно цитируемому исследованию, наследуют не только достоинства, но и недостатки других классов систем, например, проблему автоматизации трансформации сложно структурированного вопроса в унифицированную и простую форму. Подход, представленный в настоящей статье, можно отнести к этому типу, а проблема автоматизации трансформации решается нами в рамках задачи сегментации.

Автор статьи [7] предложил использовать так называемые метарегулярные выражения для отражения вариантов структуры ответа. В этой статье приведены алгоритм проверки ответов на вопросы типа «Определение» и способ автоматической генерации метарегулярных выражений для правильных ответов. По результатам тестирования разработанного прототипа автор заключает, что представленный им подход достаточно эффективно осуществляет проверку ответов на вопросы типа «Определение» для заданной предметной области. Представленные метарегулярные выражения расширяют язык обычных регулярных выражений путем добавления специальных метасимволов, которые могут быть заменены другими метарегулярными выражениями, реализуя таким образом возможность рекурсивной обработки сегментов текста. Знаки пунктуации и такие лексемы, как «это», «который», «если», выделяются в отдельную категорию, служащую точками опоры для разделения текста ответа на сегменты, соответствующие метасимволам.

В [8] представлена разработка программной системы, способной проходить тест 2015 English Entrance Exams. Системе даются текст и набор связанных с ним вопросов с вариантами ответа, на которые она дает ответ путем обработки текста. Стоящая здесь задача не связана с естественно-языковой обработкой ответа, однако авторами предложен интересный подход к семантической сегментации текста. Сегментация состоит из двух этапов: разделение текста на абзацы и сравнение текста вопроса с каждым абзацем путем использования алгоритма C99 для нахождения параграфа, к которому относится вопрос. Далее для выбора правильного ответа система сравнивает каждый вариант ответа с выбранным абзацем, используя модель, обученную на основе опорных векторов.

Авторы статьи [9] представили систему, использующую специальную метрику GAN-LCS для оценки ответов на основе их близости к эталонному ответу. Эта метрика предполагает использование грамматики связей в качестве основной модели. Описаны подходы к автоматической генерации эталонных ответов при создании контрольных тестов преподавателем на основе метода Maximum Marginal Relevance, что облегчает заполнение модели ответа. При этом система не использует машинное обучение, а представленная метрика, согласно исследованию авторов, достаточно эффективна. Однако в этой системе использован заранее заданный корпус текстов, что затрудняет ее применение в областях, где достаточное количество текстов недоступно. При этом отсутствует проблема сегментации, так как метрика близости ответа к эталонному вычисляется целиком для всего текста ответа.

В [10] представлен подход к разработке автоматизированной системы контроля знаний на основе онтологии предметной области. Разработанный алгоритм производит сегментацию текста, выявление ключевых слов и сравнение полученного набора слов с ключевыми словами эталонных ответов с использованием метода латентно-семантического анализа TF-IDF. В данном случае семантический анализ сведен к статистическому, а использованный авторами алгоритм сегментации основан на языковых особенностях китайского языка, для которого и была разработана система, поэтому нельзя считать данный подход универсальным.

В [11] описана система автоматической проверки ответов на открытые вопросы на русском языке. В качестве лингвистического процессора в системе использован Томита-парсер от компании Яндекс, который поддерживает токенизацию, морфологический анализ и выделение сущностей. Для обработки естественного языка данный инструмент использует заданные пользователем грамматики и словарь сущностей.

На основе проведенного обзора нами сделан вывод о высоком потенциале использования подхода к обработке естественно-языковых ответов, основанного на таких правилах, как синтаксические шаблоны, регулярные выражения и «мешок слов». Однако такой подход не был в достаточной степени развит для обработки ответов, формулируемых обучаемым в свободной форме, в частности, из-за трудоемкости задачи настройки правил, трудностей при верификации построенных алгоритмов, нетривиальной реализации сегментации текста ответа. Решение ряда из названных проблем с учетом опыта исследователей, упомянутых в обзоре, получено в настоящей статье.

3. Разработка прототипа

На основе представленной теоретической модели нами был разработан экспериментальный прототип с использованием языка программирования Python и веб-фреймворка Django для графического интерфейса. Этот прототип детально описан в [2]. Дальнейшая разработка прототипа в рамках проведенного исследования включала реализацию грамматик ИКГ4-ОПИСАНИЕ и ИКГ4-ОПРЕДЕЛЕНИЕ для соответствующих типов вопросов, а также поддержку рекурсивной сегментации текста ответа.

Структура ИКГ4-ОПИСАНИЕ представлена на рис. 2. Данная грамматика является рекурсивной структурой, состоящей из линейных сегментов-последовательностей, соответствующих ИКГ2 (для типовых отношений) и ИКГ3 (для составных отношений), поскольку ответ на вопрос типа «Описание» предполагает раскрытие типовых и составных отношений, связанных с главным понятием вопроса. Структура ИКГ4-ОПРЕДЕЛЕНИЕ представлена на рис. 3. Она включает ИКГ4-ОПИСАНИЕ в качестве рекурсивной компоненты и расширяет ожидаемую структуру ответа концептулой SS – главным понятием вопроса и SS_{OP} – обобщенным понятием, которое стоит на более высоком уровне в иерархической структуре предметной области над главным понятием (гипероним главного понятия). Таким образом, ответ на вопрос типа «Определение» предполагает включение не только отношений, связанных с главным концептом, но и показатель его иерархического отношения. В рамках экспериментальной апробации прототипа системы оценки ответов был разработан модуль тестирования вопросов, позволяющий оценить работу алгоритма для введенных ответов. При введении ответа форма отображает: канонизированное представление ответа, вектор ситуации, списки неопределенных лексем и запрещенных лексем, извлеченных из ответа.

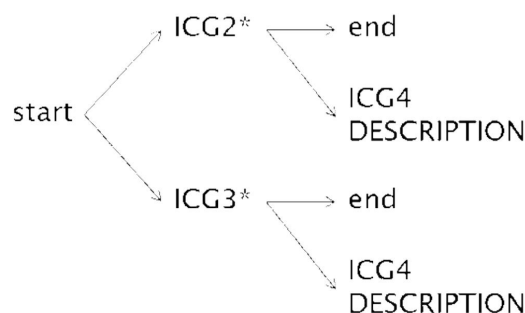


Рис. 2. Структура ИКГ4-ОПИСАНИЕ

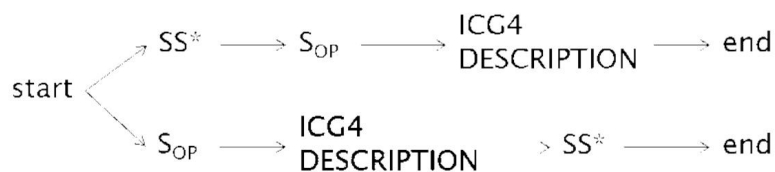


Рис. 3. Структура ИКГ4-ОПРЕДЕЛЕНИЕ

4. Результаты эксперимента

Для оценки разработанных моделей в рамках проведенного эксперимента были подготовлены вопросы и модели ответов типа «Определение» и «Описание». Пример модели ответа для одного из вопросов набора представлен в табл. 1. В эксперименте участвовали студенты бакалавриата Института вычислительной математики и информационных технологий Казанского федерального университета, обучающиеся по направлению «Прикладная информатика». Для аутентичности исследования студенты не были проинформированы об эксперименте и проходили вопросно-ответный контроль знаний в полном соответствии со стандартными правилами ручной проверки ответов преподавателем. В результате были получены и обработаны алгоритмом 20 ответов студентов.

Табл. 1

Пример модели ответа

Вопрос	Что такое компилятор?
Модель ответа (без запрещенных, необязательных и интервальных лексем)	
<i>SS</i>	компилятор
<i>SO</i>	лексическ&+анализ&, лексическ&, синтаксическ&+анализ&, синтаксическ&, этап&+генерац&, этап&+генерац&+код&, генерац&+код&, генерац&, этап&+оптимизац&+код&, этап&+оптимизац&, оптимизац&+код&, оптимизац&, этап&+загрузк&, загрузк&, состав+операционн&+систем&, операционн&+систем&, библиотек&+систем&, систем&+библиотек&
<i>SOP</i>	набор+программ&, программ&+систем&, программ&, систем&
<i>R_{SO}</i>	состоит+из, включает, предшествует, входит+в, находится+в, реализует
<i>R_{OS}</i>	состоит+из, включает, входит+в, выполняет&+после, реализу*тся+в
<i>SA</i>	текст+на+языке+высокого+уровня, язык*+высокого+уровня, текст+на+яву, яву
<i>SP</i>	текст+в+машинн&+код&, текст+на+языке+машинн&+код&, язык+машинн&+код&, машинный+код, текст+на+ямк, ямк
<i>R_A</i>	перевод&
<i>R_P</i>	получа&

На первом этапе эксперимента оценка алгоритма была проведена с использованием подготовленных эталонных ответов, которые точно должны были быть полностью разобраны алгоритмом. Результаты полученной оценки представлены в табл. 2. Согласно этим результатам, лексический процессор корректно и полно проанализировал естественно-языковой текст ответов, получив канонизированное

представление ответов без нераспознанных лексем. Семантический интерпретатор, в данном случае с учетом значения S_6 вектора ситуации, отметил эталонный ответ как полностью соответствующий ИКГ. Таким образом, можно заключить, что прототип работает успешно.

Табл. 2

Пример оценки эталонного ответа

Ответ	Это программа, которая состоит из лексического и синтаксического анализатора, включает также этапы генерации и оптимизации кодов. Переводит ЯВУ в ЯМК и предшествует этапу загрузки. Компилятор входит в состав операционной системы и находится в библиотеке системы.
Канон. ответ	LN(Это) – SOP(программа) – LN(которая) – RSO(состоит из) – SO(лексического) – LN(и) – SO(синтаксического анализатора) – RSO(включает) – LN(также) – SO(этапы генерации) – LN(и) – SO(оптимизации кодов) – RA(Переводит) – SA(ЯВУ) – GPP(в) – SP(ЯМК) – LN(и) – RSO(предшествует) – SO(этапу загрузки) – SS(Компилятор) – RSO(входит в) – SO(состав операционной системы) – LN(и) – RSO(находится в) – SO(библиотеке системы)
Вектор ситуации	$S_1 = 1.4$; $S_2 = 0$; $S_3 = 0$; $S_4 = 1$; $S_5 = 0$; $S_6 = 0$; $S_7 = 2$

На втором этапе алгоритмом были обработаны ответы, полученные от студентов. В результате в среднем только половина из них была корректно разобрана и оценена. Ответы имели нераспознанные лексемы, не учтенные в модели ответа, что приводило в некорректному разбору. По итогам эксперимента выявлены следующие группы ситуаций некорректного функционирования алгоритма и пути их разрешения:

1. Весь ответ, согласно полученному вектору ситуации, отмечен как неправильный в случае, когда его часть не была распознана алгоритмом. В данной группе ситуаций некорректного функционирования выявляется проблема реализации процедуры сегментации, требующей завершения сегментов текста только распознаваемыми лексемами для корректного нахождения данных сегментов. Для решения данной группы ситуаций предложено использование контекстно-свободного алгоритма разбора ответа, при котором каноническое представление ответа разбивается на сегменты не только по цепочке слева направо, но и справа налево. В результате такие ответы должны иметь оценку $S_6=1$ в векторе ситуации, которая соответствует частичной правильности ответа.
2. Формулировка ответов в виде определения для вопросов типа «Описание». Оба типа вопросов имеют похожие ИКГ, однако ИКГ4 ОПРЕДЕЛЕНИЕ включает в себя также требование использования в ответе гиперонима для главного понятия вопроса. Данная группа ситуаций влечет некорректное распознавание соответствующей концептулы S_{OP} и в соответствии с ситуацией 1 ведет к отметке ответа целиком как неправильного. Разрешение ситуаций этой группы требует реализации процедуры автоматического определения ИКГ ответа для близких между собой грамматик.

Заключение

Дальнейшее совершенствование алгоритма обработки ответов будет направлено на разрешение выявленных ситуаций некорректного функционирования. Ответы студентов, полученные в рамках эксперимента, будут использованы и в следующих экспериментах. Но уже по результатам проведенного исследования можно заключить, что прагматически-ориентированный подход может быть применен для автоматической обработки ответов на сложные вопросы типа «Определение» и «Описание», а предложенный метод семантико-синтаксической рекурсивной сегментации, при всех его недостатках, позволяет свести решаемую задачу к обработке простых сегментов вопроса с использованием уже реализованных грамматик. Другое направление будущих работ – это развитие прототипа путем автоматизации создания модели ответа, поскольку для преподавателя это наиболее трудоемкая задача при создании контрольного теста. Разработки, уже проведенные в данном направлении, включают использование онтологий предметной области для автоматизированной генерации опросов.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Литература

1. *Khasianov A., Suleymanov D., Marchenko A.* Three agent platform approach for digital education environment // INTED2017 Proc. 2017. P. 8580–8587. <https://doi.org/10.21125/inted.2017.2031>.
2. *Suleymanov D., Prokopyev N.* Development of prototype of natural language answer processor for e-learning // Artificial Intelligence (RCAI 2020): Proc. 18th Russ. Conf. Ser.: Lecture Notes in Computer Science. V. 12412. Cham: Springer, 2020. P. 448–459. https://doi.org/10.1007/978-3-030-59535-7_33.
3. *Suleymanov Dzh., Prokopyev N.* Experimental prototype of the pragmatically oriented e-assessment algorithm for automatic natural language answer grading // EDULEARN20 Proc. 2020. P. 1917–1924. <https://doi.org/10.21125/edulearn.2020.0617>.
4. *Сулейманов Д.Ш.* Система семантического анализа ответных текстов обучаемого на естественном языке // Онтол. проект. 2014. № 1 (11). С. 65–77.
5. *Burrows S., Gurevych I., Stein B.* The eras and trends of automatic short answer grading // Int. J. Artif. Intell. Educ. 2015. V. 25. P. 60–117. <https://doi.org/10.1007/s40593-014-0026-8>.
6. *Мишунин О.Б., Савинов А.П., Фирстов Д.И.* Состояние и уровень разработок систем автоматической оценки свободных ответов на естественном языке // Соврем. наукоемк. технол. Техн. науки. 2016. № 1. С. 38–44.
7. *Мерзляков Д.А.* Генерация регулярных выражений для автоматизации проверки тестов открытого характера // Студенч. научн. форум – 2013: V Междунар. студенч. науч. конф. 2013.
8. *Ziai R., Rudzewitz B.* CoMiC: Exploring text segmentation and similarity in the English entrance exams task // CLEF2015 Working Notes / Cappellato L., Ferro N., Jones G.J.F., Juan E.S. (Eds.). Ser.: CEUR Workshop Proceedings. 2015. V. 1391.
9. *Pribadi F.S., Permanasari A.E., Adji T.B.* Short answer scoring system using automatic reference answer generation and geometric average normalized-longest common subsequence (GAN-LCS) // Educ. Inf. Technol. 2018. V. 23, No 6. P. 2855–2866. <https://doi.org/10.1007/s10639-018-9745-z>.
10. *Xu L.-X., Wang N., Xu L., Li L.-Y.* Research of automated assessment of subjective tests based on domain ontology // Multidisciplinary Social Networks Research (MISNC 2014):

Proc. Int. Conf. Ser.: Communications in Computer and Information Science. V. 473. Berlin, Heidelberg: Springer, 2014. P. 445–453.
https://doi.org/10.1007/978-3-662-45071-0_37.

11. *Кожевников В.А., Сабинин О.Ю.* Система автоматической проверки ответов на открытые вопросы на русском языке // Науч.-техн. ведом. СПбГПУ. Информ. Телекомм. Управл. 2018. Т. 11, № 3. С. 57–72. <https://doi.org/10.18721/JCSTCS.11306>.

Поступила в редакцию 24.12.2023

Принята к публикации 11.10.2024

Прокопьев Николай Аркадиевич, старший преподаватель Института вычислительной математики и информационных технологий

Казанский (Приволжский) федеральный университет

ул. Кремлевская, д. 18, г. Казань, 420008, Россия

E-mail: nikolai.prokopyev@gmail.com

ISSN 2541–7746 (Print)

ISSN 2500–2198 (Online)

UCHENYE ZAPISKI KAZANSKOGO UNIVERSITETA.
SERIYA FIZIKO-MATEMATICHESKIE NAUKI
(Proceedings of Kazan University. Physics and Mathematics Series)

2024, vol. 166, no. 4, pp. 580–593

ORIGINAL ARTICLE

doi: 10.26907/2541-7746.2024.4.580-593

**Automatic Grading of Answers in Knowledge Control
for “Definition” and “Description” Question Types**

N.A. Prokopyev

Kazan Federal University, Kazan, 420008 Russia

E-mail: nikolai.prokopyev@gmail.com

Received December 24, 2023; Accepted October 11, 2024

Abstract

The progress in developing an effective automatic knowledge control system is directly associated with creating and implementing a software module for grading answers to test questions formulated in natural language. Previously, an experimental prototype of such a system was designed, and a study was performed where short answers to basic question types provided by students were examined via a pragmatically oriented question-answer text processing algorithm, considering its outputs and exposing flaws. This article introduces the next iteration of the algorithm tailored to handle more complex question types that requires the identification of relations such as “Definition” and “Description.” The key features of the

enhanced algorithm were outlined, with a particular focus on the problem of segmenting answers into meaningful chunks, a task for which processing methods have already been found. The results of an experiment based on the developed prototype with obtaining answers from students and a thorough analysis of the instances of the system's incorrect behavior were discussed.

Keywords: natural language processing, automatic answer grading, e-learning

Conflicts of Interest. The authors declare no conflicts of interest.

Figure Captions

Fig. 1. Answer processing algorithm.

Fig. 2. ICG4-DESCRIPTION structure.

Fig. 3. ICG4-DEFINITION structure.

References

1. Khasianov A., Suleymanov D., Marchenko A. Three agent platform approach for digital education environment. *INTED2017 Proc.*, 2017, pp. 8580–8587. <https://doi.org/10.21125/inted.2017.2031>.
2. Suleymanov D., Prokopyev N. Development of prototype of natural language answer processor for e-learning. *Artificial Intelligence (RCAI 2020): Proc. 18th Russ. Conf. Ser.: Lecture Notes in Computer Science*. Vol. 12412. Cham, Springer, 2020, pp. 448–459. https://doi.org/10.1007/978-3-030-59535-7_33.
3. Suleymanov Dzh., Prokopyev N. Experimental prototype of the pragmatically oriented e-assessment algorithm for automatic natural language answer grading. *EDULEARN20 Proc.*, 2020, pp. 1917–1924. <https://doi.org/10.21125/edulearn.2020.0617>.
4. Suleymanov D.Sh. A system for semantic processing of natural language responses. *Ontol. Proekt.*, 2014, no. 1 (11), pp. 65–77. (In Russian)
5. Burrows S., Gurevych I., Stein B. The eras and trends of automatic short answer grading. *Int. J. Artif. Intell. Educ.*, 2015, vol. 25, pp. 60–117. <https://doi.org/10.1007/s40593-014-0026-8>.
6. Mishunin O.B., Savinov A.P., Firstov D.I. Current progress and development trends of systems for automatic grading of free answers in natural language. *Sovrem. Naukoemkie Tekhnol. Tekh. Nauki*, 2016, no. 1, pp. 38–44. (In Russian)
7. Merzlyakov D.A. Generation of regular expressions for automatic grading of open-ended tests. *Studencheskii nauchnyi forum – 2013: V Mezhdunar. studench. nauch. konf.* [Student Research Forum – 2013: Proc. V Int. Student Sci. Conf.], 2013. (In Russian)
8. Ziai R., Rudzewitz B. CoMiC: Exploring text segmentation and similarity in the English entrance exams task. *CLEF2015 Working Notes*. Cappellato L., Ferro N., Jones G.J.F., Juan E.S. (Eds.). Ser.: CEUR Workshop Proceedings, 2015, vol. 1391.
9. Pribadi F.S., Permanasari A.E., Adji T.B. Short answer scoring system using automatic reference answer generation and geometric average normalized-longest common subsequence (GAN-LCS). *Educ. Inf. Technol.*, 2018, vol. 23, no. 6, pp. 2855–2866. <https://doi.org/10.1007/s10639-018-9745-z>.
10. Xu L.-X., Wang N., Xu L., Li L.-Y. Research of automated assessment of subjective tests based on domain ontology. *Multidisciplinary Social Networks Research (MISNC 2014): Proc. Int. Conf. Ser.: Communications in Computer and Information Science*. Vol. 473. Berlin, Heidelberg, Springer, 2014, pp. 445–453. https://doi.org/10.1007/978-3-662-45071-0_37.

-
11. Kozhevnikov V.A., Sabinin O.Yu. Automatic system for processing answers to open-ended questions in Russian. *Inf. Telekommun. Upr.*, 2018, vol. 11, no. 3, pp. 57–72.
<https://doi.org/10.18721/JCSTCS.11306>. (In Russian)
-

Для цитирования: Прокопьев Н.А. Автоматизированная оценка ответов при контроле знаний для вопросов типа «Определение» и «Описание» // Учен. зап. Казан. ун-та. Сер. Физ.-матем. науки. 2024. Т. 166, кн. 4. С. 580–593.
<https://doi.org/10.26907/2541-7746.2024.4.580-593>.

For citation: Prokopyev N.A. Automatic grading of answers in knowledge control for “Definition” and “Description” question types. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki*, 2024, vol. 166, no. 4, pp. 580–593.
<https://doi.org/10.26907/2541-7746.2024.4.580-593>. (In Russian)