

## О ПРОБЛЕМЕ АВТОМАТИЗАЦИИ ПРОВЕРКИ ОТКРЫТЫХ ВОПРОСОВ ПРИ ТЕСТИРОВАНИИ

*Тимошек Е.С., студент гр. 151001*

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Парамонов А.И. – канд. техн. наук, доцент*

**Аннотация.** В статье затронута проблема обработки естественных языков в контексте нахождения семантической близости между двумя фрагментами текста, один из которых является эталонным ответом на вопрос, а другой является ответом на тот же вопрос, который был дан тестируемым. Рассмотрены основные сложности, которые возникают в процессе семантического сравнения различных текстов. Приведено проектное решение обозначенных вопросов в виде программного средства.

**Ключевые слова.** Обработка текста, контекст, семантическое сравнение, Word2vec, Bert, проверка тестов, автоматизация.

С каждым годом все больше сфер жизни человека подвергается цифровизации. В настоящее время вычислительные машины используются повсеместно: от бухгалтерского учета, до медицинских исследований. Главное преимущество при использовании вычислительной техники заключается в возможности делегировать часть рутинных задач, не требующих творческих навыков на технику. Это позволяет автоматизировать многие процессы в самых разных отраслях, что повышает продуктивность и качество выполняемой работы. На данный момент современные информационно-коммуникационные технологии активно используются в образовательном процессе. Существует множество программных средств и платформ для повышения качества учебного процесса, автоматизации отдельных процессов и в целом цифровизации [1]. Одним из аспектов образовательного процесса является проверка знаний (различные виды контроля). Контроль знаний учащихся в форме тестирования сегодня нашел широкое применение. Используются системы автоматического тестирования учащихся, которые позволяют создавать тесты с различными видами вопросов: вопросы с вариантами ответа, вопросы с пропущенным словом, вопросы, где нужно ввести короткий числовой ответ, вопросы на сопоставление и многие другие. Автоматическая проверка данных вопросов не вызывает затруднений, так как заранее известно каким должен быть правильный ответ либо, в некоторых случаях, варианты правильного ответа. Что же касается открытых вопросов, где учащемуся необходимо порассуждать на тему и дать собственную формулировку, то самые популярные системы автоматической проверки ответов не поддерживают автоматическую проверку таких вопросов.

Открытым вопрос в данном контексте – это вопрос, требующий от учащегося передать суть некоторого явления своими словами либо требующий порассуждать на некоторую тему, выдвинуть свои гипотезы и предположения. Как правило, ответ на подобный вопрос включает в себя от нескольких предложений до нескольких абзацев текста. Ключевое отличие данного вида вопросов при автоматической проверке от перечисленных выше видов заключается в том, что верный ответ может иметь огромное количество вариаций, которые могут отличаться как порядком слов и очередностью утверждений, так и иметь различные способы раскрытия сути одного и того же явления, примером чего можно считать приведение различных аналогий. Очевидно, что предусмотреть все возможные варианты верного ответа невозможно, требуется сравнивать семантический смысл, вложенный в текст, являющийся эталоном и в текст, который необходимо проверить. Так как по озвученным выше причинам они не могут совпадать полностью, то необходимо проверить укладывается ли проверяемый ответ в некоторый интервал, в котором ответ считается верным. Интервал можно установить свой для каждого вопроса в зависимости от того, ожидается ли точный ответ на поставленный вопрос от учащегося или достаточно его мыслей в верном направлении.

При анализе естественных языков необходимо учитывать специфику конкретного языка. В мире существует множество языков, каждый из которых может иметь свои уникальные правила и особенности. Это может выражаться как в способе писать слова, как в языках на подобие китайского, где каждый иероглиф сам по себе имеет смысл и в некотором роде является словом, так и в способе построения предложений. Не стоит забывать и про то, что на смысл, вкладываемый в текст, написанный на определенном языке, влияет культурная среда, в которой живет человек. Кроме того, перевод текста с одного языка на другой ведет к потере части вложенного в него смысла, а устойчивые выражения, используемые в одном языке, могут звучать как не имеющий смысла набор слов в другом, или просто не иметь подходящего аналога в этом языке, но сам смысл высказывания понятен.

Русский язык, как и другие языки, имеет свои особенности. Одной из таких особенностей является важность пунктуации. В одном предложении может быть неограниченное количество знаков пунктуации, которые используются для связи частей предложения. Данная проблема возникает из-за

того, что предложения в русском языке есть сложносочиненные предложения, которые состоят из двух равноправных частей, и сложносочиненные предложения, в которых одна часть зависит от другой. При этом каждая часть предложения может состоять из произвольного количества частей поменьше, использующих те же способы связи. В данных обстоятельствах знаки пунктуации указывают на то, как и к какой конкретно части предложения относится другая часть. Кроме того, в русском языке предложение, являющееся вопросительным, может отличаться от утвердительного предложения исключительно интонацией в устной речи и знаком точки или вопроса в письменном виде. Для примера можно привести предложение «Он уже ушел?». Если в конце предложения стоит точка, то один собеседник передает другому информацию о том, что кто-то уже ушел. Если же в конце предложения будет вопросительный знак, то один собеседник запрашивает информацию у другого, ожидая получить от него ответ. Получается два разных предложения, состоящих из одних и тех же слов в одной и той же последовательности, но несущих в себе различный смысл. В качестве примера влияния запятой на смысл предложения можно привести обращения, которые отделяются запятыми от остального предложения и служат указанием, кто является адресатом предложения. Для примера можно привести предложение «Сергей, пошел бы в лес». В случае, когда запятая присутствует, один человек обращается к другому и предлагает ему сходить в лес. Если убрать запятую, то получится, что один собеседник сообщает другому собеседнику информацию о том, что Сергей с наибольшей вероятностью отправился бы в лес. Так же возможны ситуации, когда запятая была пропущена. В таком случае определить смысл некоторый предложений становится проблематично, для примера известная фраза «казнить нельзя помиловать» в зависимости от запятой приобретает прямо противоположный смысл: в случае, когда запятая перед «нельзя», фраза означает окончательный приговор о смертной казни, который не подлежит обжалованию, в то время как запятая после «нельзя» превращает фразу в помилование. Для того, чтобы определить точный смысл предложения в случае, когда запятая пропущена, необходимо учитывать контекст, но даже так есть шанс, что определить точно не получится, и тогда придется проверять несколько вариантов расстановки знаков пунктуации.

Еще одной проблемой при обработке русского языка являются синонимы и омонимы, то есть близкие по смыслу слова, но совершенно разные по написанию, и одинаковые по написанию слова, но разные по смыслу. Проблема с синонимами стоит менее остро, так как существует опция все синонимичные слова условно считать одним и тем же словом. Что касается омонимов, то единственный способ установить точное значение слова-омонима заключается в анализе контекста, причем имеется в виду контекст не только предложения, но и всего текста целиком. Например, в предложении «Он смотрит на лук.» невозможно однозначно определить, идет речь об овоще или об оружии. Но если это происходит в огороде, то вероятнее всего речь идет об овоще. Наиболее точно это определяется лишь тогда, когда описываются действия с предметом, который обозначается этим словом, возможные для одного значения слова и невозможные для другого значения, либо по каким-то характеристикам, имеющим смысл только для одного из значений, либо по каким-то связанным словам, к примеру, если речь пойдет про тетиву лука, то это однозначно оружие. Однако, проблему с омонимами можно считать несущественной в тех случаях, когда сравниваются текстовые фрагменты на схожую тему, так как у них будет общий контекст, а значит и вероятность, что смысл слов-омонимов будет одинаковым, будет близка к единице. Современные модели, такие как Word2vec и Bert в своей работе учитывают контекст слова, для определения его семантического значения [2].

На ряду со словами-омонимами, проблемой также являются опечатки, которые неизбежно допускаются людьми. В общем случае для исправления опечаток используются различные способы, опирающиеся на метрики на подобие расстояния Левенштейна [3]. В сравнении с английским языком, в русском языке куда меньше слов, которые различаются на одну или две буквы, либо имеют две буквы, поменянные местами. Это уменьшает возможные варианты того, что предполагалось написать изначально. Кроме того, это уменьшает количество ситуаций, когда в слове была допущена опечатка и получилось другое существующее слово. Для примера можно привести английские слова beer (пиво) и bear (медведь) или русские пень и тень, однако в русском языке из-за наличия множества форм слов отличить одно слово от другого становится проще, так как они могут склоняться по-разному. В целом, проблема с опечатками может быть решена схожими средствами, что и проблема со словами-омонимами, с добавлением использования расстояния редактирования для того, чтобы определять наиболее вероятные варианты правильного написания слова с учетом количества ошибок, которое необходимо было допустить в слове, чтобы оно получилось таким.

Стоит отметить, что порядок слов в русском языке может меняться в широких пределах. В русском языке есть несколько определенных общепринятых способов построения предложений, однако любой человек может использовать свой уникальный стиль общения, который будет немного отличаться. Более того, в некоторых ситуациях смысл предложения может никак не поменяться, а в некоторых порядок слов влияет на смысл предложения. В качестве примера можно привести «дом стоит уже двадцать лет» и «дом стоит уже лет двадцать» – в первом случае сообщается точная информация, что дом стоит именно двадцать лет, во втором случае информация приблизительная, и на самом деле дом может стоять и девятнадцать, и двадцать три, и двадцать лет. Кроме того, если

рассматривать сложносочиненные или сложноподчиненные предложения, то обмен местами подлежащих или сказуемых из разных частей предложения исказит его смысл значительно. В рамках одной части предложения перестановка слов местами влияет на итоговый смысл предложения в меньшей степени, при перестановке однородных членов предложения эффекта и вовсе не будет, однако перестановка слов может сменить отношения между словами, к примеру, перемешав прилагательные разных предметов. Таким образом в русском языке необходимо учитывать порядок слов предложения для корректного семантического сравнения различных текстовых фрагментов.

Еще одна сложность, которая возникает в процессе семантического анализа текстового фрагмента, заключается в местоимениях, таких как «он», «она», «они» и другие. Сами по себе местоимения не несут полной информации, сообщая лишь о числе некоторых субъектов и роде субъекта, если он один. Однако, местоимения используются в предложениях для того, чтобы заменять собой существительные, которые уже были ранее упомянуты. Это делается для того, чтобы избежать повторяемости слов, которая делает текст менее выразительным и топорным. Например, в тексте «Ваня встал утром, после чего он пошел чистить зубы», имя Вани заменяется местоимением «он». Это обстоятельство вынуждает учитывать, что было написано до встреченного местоимения, чтобы иметь возможность определить, какое слово было заменено местоимением. Стоит упомянуть, что не всегда местоимение относится к ближайшему слову, которому оно предшествует. Например, «Ученики сдали экзамены, чему они несказанно рады». В данном предложении местоимение «они» однозначно относится к ученикам, а не к экзаменам, хотя оба слова имеют множественное число и могут быть заменены словом они. Кроме того, бывают ситуации, где однозначно определить, к какому слову относится местоимение не представляется возможным. Примером этого может послужить предложение «Василий увидел Максима, и он улыбнулся». В данном случае улыбнуться мог как Василий, так и Максим. В данной ситуации разрешить неоднозначную ситуацию может помочь контекст предыдущих предложений, либо последующие предложения. То есть, если после этого предложения будет написано что-то вроде «Он уже давно не видел Максима улыбающимся», то очевидно, что речь шла о Максиме. Такая ситуация, когда возникает неоднозначность связи между местоимением и заменяемым им словом, которую невозможно разрешить без анализа предшествующих и последующих предложений, называется анафорой. Для разрешения анафоры могут быть использованы различные методы, в том числе те, которые основаны на машинном обучении [4].

Еще одной значительной проблемой при выделении смыслового наполнения текста являются фразеологизмы, то есть устойчивые словосочетания, а также крылатые выражения. Основная проблема заключается в том, что несколько слов, составляющих фразеологизм имеют совершенно иной смысл, чем их комбинация в фразеологизме. Например, словосочетание «повесить нос» означает расстроиться. Буквальное трактование данного словосочетания предполагает, что кто-то повесил куда-то нос. Помимо фразеологизмов, существуют так же крылатые выражения, которые по сути своей являются теми же фразеологизмами. Ключевое отличие между ними заключается в том, что фразеологизмы возникают в какой-то конкретной культуре и не понятны для людей из других культур. В свою очередь крылатые выражения связаны с историческими событиями и личностями, с мифами или известными литературными произведениями и авторами. Обработка фразеологизмов и крылатых выражений может с одной стороны может быть выполнена посредством предобработки текста, во время которой все устойчивые словосочетания будут заменены их истинным значением, так как в большинстве случаев слова, составляющие фразеологизм неразрывно связаны и идут в строго определенном порядке, однако, несмотря на то, что многие фразеологизмы всегда являются фразеологизмами, некоторые из них могут быть использованы в буквальном значении. Например, фразеологизм «серая мышь», который означает скромного и неприметного человека, может оказаться не фразеологизмом в случае, речь идет мышах и говоря про серую мышь человек действительно будет иметь в виду мышь серого цвета. Как и в случае с омонимами, необходимо анализировать контекст, в котором устойчивое словосочетание используется, что может быть сделано при помощи различных способов, например, используя метод Bag-of-ngrams или GloVe [5].

Проанализировав существующие сложности и наиболее популярные методы и алгоритмы был сделан вывод, что разработка систем для проверки открытых вопросов должна включать использование технологий машинного обучения и обработки естественного языка для создания моделей, способных оценивать содержание открытых ответов по семантическому смыслу. Важно учитывать языковые особенности, создавая специализированные модели для различных языков, которые будут учитывать их уникальные правила и грамматику. Кроме того, необходимо интегрировать анализ пунктуации, разрабатывая алгоритмы, которые помогут определить структуру и смысл предложений. Для обработки синонимов и омонимов можно использовать известные контекстуальные модели на базе трансформеров, такие как BERT, что позволит лучше использовать контекст. Коррекция опечаток требует разработки алгоритмов, учитывающих морфологические особенности языка, чтобы обеспечить более точную коррекцию. Анализ порядка слов является еще одной важной задачей; для этого следует использовать модели, которые могут учитывать изменения порядка слов и их влияние на смысл, возможно, с использованием глубоких нейронных сетей.

При разработке алгоритмов были изучены существующие подходы сравнения текстовых фрагментов, их эффективность и особенности применения. Данные подходы включают в себя: посимвольное сравнение текстов, расстояние Левенштейна, расстояние Дамерау-Левенштейна, Bag-of-words, метрика Жаккара, TF-IDF, word2vec, Bert, GloVe, Doc2vec и косинусное сходство. Наиболее подходящими для быстрого предварительного сравнения оказались алгоритмы, основанные на расстоянии Левенштейна, для выявления текстов, которые практически полностью совпадают с ожидаемым ответом, и алгоритмы, основанные на метрике Жаккара, позволяющие получить схожесть двух текстов основываясь на наборе слов, входящих в них. Для повышения точности может быть использованы модификации на подобие TF-IDF для уменьшения веса слов, не несущих в себе значения (предлоги и союзы и другие). Основное преимущество указанных алгоритмов является скорость, так как они довольно простые и не требуют вычислительных ресурсов, однако это негативно сказывается на точности оценки, так как не учитывается контекст, порядок слов в предложениях, их структура и последовательность, а также не учитывает фразеологизмы и омонимы. Кроме того, слова используются в своей начальной форме, что также несет потерю смысла. Более точную оценку дают алгоритмы, использующие косинусное сходство. Для данного метода необходимо некоторым способом преобразовать сравниваемые фрагменты текста в векторное пространство, в котором они будут представлять свое смысловое значение. Точность данного метода зависит от способа перевода текста в векторное пространство. Наиболее эффективными оказались модели Bert и Word2vec, а также модель Doc2vec, основанная на Word2vec. Преимущества Word2vec заключается в скорости обучения и в возможности учитывать контекст слов в предложении или нескольких предложениях. Однако, данная модель преобразует лишь одно слово в векторное пространство, что требует дополнительных алгоритмов для перевода векторов слов в тексте в один вектор, который будет представлять собой вектор текста. Для этого используется усредненный вектор всех слов, входящих в текст, при этом использование TF-IDF в данном случае не дает ощутимого прироста точности оценки. Либо может быть использована модификация Word2vec под названием Doc2vec, которая преобразует весь текст целиком в векторное пространство. Данная модель демонстрирует большую точность, чем модель Word2vec, при том, что может быть обучена на том же наборе данных и за то же время.

Обозначенные проблемы предлагается решить путем создания программного средства, которое будет обладать следующим функционалом: семантическое сравнение текстов (определение меры схожести двух текстов в диапазоне), функция создания тестов, которые состоят из вопросов, требующих развернутых ответов (для создания вопроса пользователь должен ввести текст самого вопроса и пример эталонного ответа, который будет считаться правильным и использован для оценки ответов учащихся), регистрация пользователей с различными ролями: ученик и преподаватель, а также процесс авторизации, чтобы обеспечить доступ к функционалу, прохождение тестов, созданных преподавателями, а также осуществление автоматической проверки всех ответов на тест и формирование отчета для преподавателя о тех работах, которые нуждаются в дополнительной проверке, так как система автоматической проверки дает оценку схожести близкую к пограничной, до которой ответ считается неверным, а после которой он считается верным.

Программное средство реализуется с использованием современных средств и технологий, включая такие языки программирования, как Java для серверной части и JavaScript для клиентской части. Для разработки серверного приложения применяется фреймворк Spring Boot, который в сочетании с Spring Data JPA обеспечивает эффективную работу с базами данных. На клиентской стороне используется React, что позволяет создавать динамичные и отзывчивые пользовательские интерфейсы. Для выполнения HTTP-запросов из React-приложения применяется библиотека Axios. Для управления зависимостями в Java используется Maven, а для JavaScript – Node.js и npm. В качестве системы управления базами данных используется PostgreSQL. Для семантического сравнения текстов в рамках программного средства используется язык Python с такими библиотеками как NLTK (Natural Language Toolkit) для токенизации, стемминга и лемматизации, spaCy, поддерживающая семантический анализ и векторизацию слов и Transformers от Hugging Face предоставляет доступ к предобученным моделям, таким как BERT, которые предоставляют инструменты для семантического сравнения текстовых фрагментов.

#### **Список использованных источников:**

1. Парамонов, А. И. Проблемы дистанционного образования и их прикладные решения в образовательных технологиях / А. И. Парамонов // Высшее техническое образование : проблемы и пути развития = Engineering education: challenges and developments : материалы X Международной научно-методической конференции, Минск, 26 ноября 2020 года / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2020. – С. 182–187.
2. Частикова, В. А. Методика обнаружения атак социальной инженерии на основе алгоритмов анализа естественного языка / В. А. Частикова, В. Г. Гуляй // Прикаспийский журнал: управление и высокие технологии. – 2022. – № 3(59). – С. 61-71.
3. Напреенко, Г. В. Программные возможности идентификации текстов: сопоставление на схожесть, установление тождества, проверка на уникальность / Г. В. Напреенко, // Вестник челябинского государственного университета, Номер 3(485). 2024. – С. 55-65.
4. Соколов, А.В. Разработка системы разрешения анафоры на основе методов машинного обучения / А. В. Соколов, Т. В. Батура // Программные продукты и системы = Software & Systems, Номер 29, 2017. – С. 461–468.

## PROCESS AUTOMATION OF CHECKING TEST QUESTIONS WITH TEXTS ANSWERS

*Timoshek E.S.*

*Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus*

*Paramonov A.I. – PhD in Technical Sciences, associate professor*

**Annotation.** The article reveals the natural language processing problem in the context of finding semantic proximity between two text fragments. One text is a reference answer to some open question, and the other considered text is a student answer to the same test question. The main difficulties that arise in the process of semantic comparison of different texts are considered. A design solution to the identified issues in the form of a software tool is provided.

**Keywords.** NLP, Text Processing, Context, Semantic Matching, Word2vec, Bert, Testing Verification, Automation processing.