

# АВТОМАТИЗАЦИЯ ПРОВЕРКИ ОРФОГРАФИИ РУССКОГО ЯЗЫКА

*Нгуен Хоанг Хуеп*

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

*Петрова Н.Е. – к.филол.н., доцент*

В статье рассматривается значение орфографии в языке. Анализируются автокоррекция правописания, преимущества и недостатки её онлайн-версий. Проводится краткий анализ некоторых онлайн-программ для исправления орфографии на русском языке.

В настоящее время информационные технологии играют роль как мягкой инфраструктуры развития, так и движущей силы всех отраслей и сфер общественной жизни. Благодаря информационным технологиям родился широкий спектр научно-производственных отраслей и новых сервисов, позволяющих решать различные задачи, преодолевать трудности в управлении и деятельности учреждений, организаций и даже целых сфер. Автоматизация – это процесс, при котором устройство или элемент автоматически управляет интеллектуальной системой машин, кроме того, он также использует искусственный интеллект для помощи в управлении устройством вместо действий человека. Самая большая польза, которую приносит автоматизация, заключается в том, что она экономит труд, усилия и время, кроме того, она используется для экономии энергии и сырья, повышения качества жизни с высокой степенью точности.

Информационные технологии являются неотъемлемой частью каждой сферы профессиональной деятельности, в том числе и лингвистики. Если буквально несколько лет назад компьютеры и соответствующее программное обеспечение не использовалось в лингвистических исследованиях или использовалось очень редко, не было обязательным даже при переводе и в обучении иностранным языкам, то на сегодняшний день преподаватели, исследователи, лингвисты нуждаются и активно используют информационные технологии в своей профессиональной среде.

На первый взгляд может показаться, что лингвистика и информатика далеки друг от друга и не пересекаются. Однако они обладают довольно тесной связью. Существует даже такое направление в прикладной лингвистике, как компьютерная лингвистика, деятельность которой обращена на использование компьютерных технологий для моделирования функционирования языка в различных ситуациях или проблемных вопросах.

В подтверждение связи между информационными технологиями и лингвистикой можно отметить, что существуют направления, находящиеся на пересечении данных сфер, и которые при этом являются уже реализованными в форме программ и различных источников в Интернете. Такие ресурсы подразделяются на классы: программы для языковой обработки и анализа текста; программы для автоматической обработки текста; программы, необходимые для преобразования текста; программы психолингвистики; программы, созданные для обработки языков и машинного перевода; вычислительные программы в лингвистике; системы, предназначенные для полнотекстового поиска.

В использовании мультимедийных технологий в лингвистике есть особенность, которая связана с тем, что пользователям можно использовать не только обучающие, но и всевозможные прикладные программы, такие как редакторы текстов, программы проверки правописания, электронные словари, электронные

таблицы, настольные издательские системы, энциклопедические и игровые программы. Постоянно появляются новые продукты, содержащие 3D графику, статистические и динамические изображения объектов, мультипликацию. Сегодня имеется возможность графического моделирования процессов, явлений, различных ситуаций. С помощью мультимедийных технологий появились способы моделирования различных сред и средств обучения [1].

Орфография – это раздел лингвистики, который связан с единообразной передачей слов и грамматических форм речи на письме [2]. Изучая иностранный язык, нужно помнить, что каждый язык имеет свою орфографию, грамматику, структуру, порядок слов и т. д. Важно изначально употреблять правильные варианты слов, поэтому при изучении иностранных языков особенно актуальны ресурсы, которые будут исправлять ошибки. Очевидно, что автоматизация проверки орфографии русского языка очень нужна иностранцам, особенно тем, кто изучает неродственные языки. Например, вьетнамцам или другим представителям азиатских стран сложно изучать русский язык, поскольку он существенно отличается от родного. Это объясняет особую актуальность различных продуктов, которые помогут вьетнамцам в изучении русского языка.

Система проверки правописания (также спел-чекер от англ. *spell checker*) – компьютерная программа, осуществляющая проверку заданного текста на наличие в нём орфографических ошибок. Найденные ошибки или опечатки отмечаются специальным образом, обычно для этого используется подчёркивание. В некоторых случаях пользователю помимо указания на места возможных ошибок предоставляется возможность выбрать один из правильных вариантов написания. Может быть также выведен комментарий, объясняющий, каким образом следует поправить текст [3].

В настоящее время существует множество онлайн-сервисов проверки русской орфографии, такие как Text.ru [4], LanguageTool [5], ADVEGO [6] и др. Мы считаем, что ими достаточно легко пользоваться. Они позволяют искать, обнаруживать, исправлять и опечатки ошибки в словах, предложениях, абзацах и даже в изображениях. На рисунках 1, 2, 3 показаны интерфейсы работы некоторых таких онлайн сервисов для проверки русской орфографии.

На рисунке 1 представлен продукт Text.ru. Это многофункциональный бесплатный сервис для обработки текстового контента. Многие подобные программы отличаются шаблонностью и скупостью интерфейса. Мы считаем, что Text.ru один из таких, однако со своими функциями он справляется. На рисунке ошибки выделены красным фоном. Проверка орфографии нашла 4 ошибки (рисунок 1):

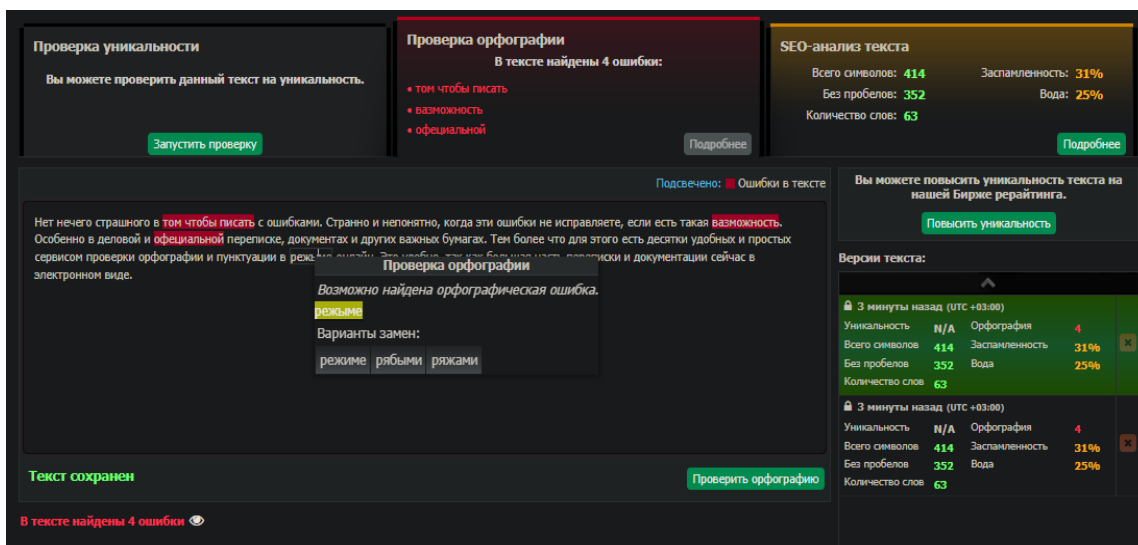


Рисунок 1 – Интерфейс онлайн-сервиса проверки русской орфографии Text.ru

На рисунке 2 представлен продукт LanguageTool. Приложение разработано для коррекции грамматики, пунктуации и стиля различных языков. Бесплатно будет дано 20 тыс. символов для проверки, если более, то услуга станет платной. Программа может находить большое количество ошибок, выделяет повтор слов и некоторые погрешности грамматики, руководствуясь правилами проверки для 25 языков, в том числе для русского. Для большинства грамматических ошибок предлагаются варианты исправления. LanguageTool можно запустить непосредственно в браузере или установить в качестве независимого приложения. Программа нашла в нашем тексте 4 ошибки (рисунок 2):

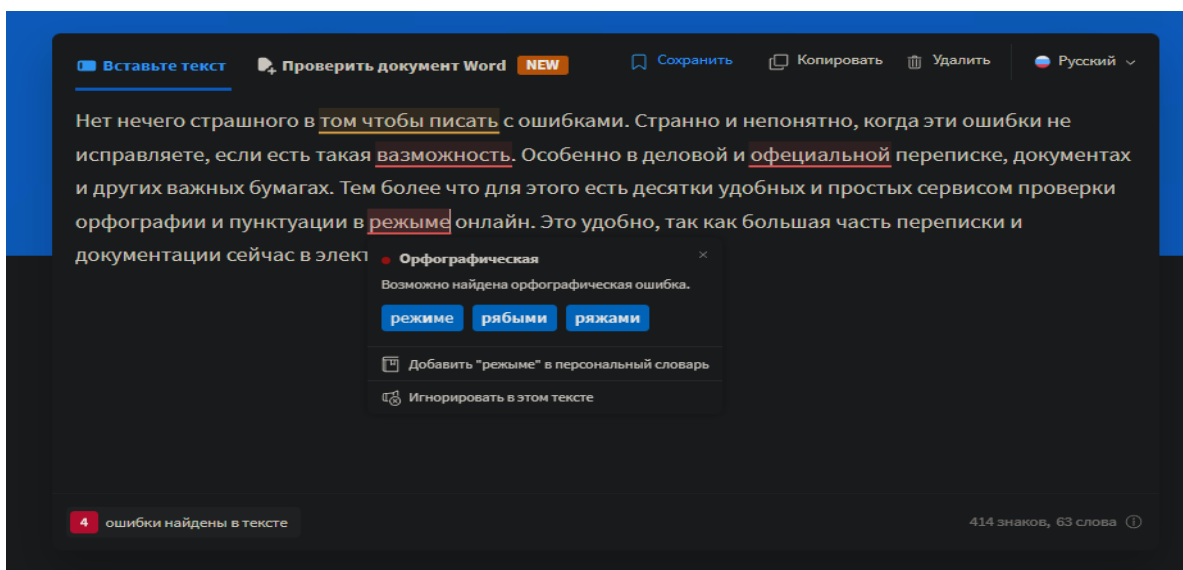


Рисунок 2 – Интерфейс работы онлайн-сервиса проверки русской орфографии LanguageTool

В свою очередь отметим, что не все ошибки были найдены. Например, в начале нашего текста написано «Нет ничего страшного». Это неправильно. Правильный вариант «Нет ничего страшного». Также не были найдены некоторые пунктуационные ошибки, например, необходима запятая в предложении «Тем более, что для этого есть ...».

На рисунке 3 представлен бесплатный сервис ADVEGO, который за один раз может проверить до 100 000 символов. Поэкспериментировав, мы убедились в том, что, если предложение не будет связано по смыслу, то ошибки в словах найти сложнее, так как программа не знает, что это за слово. Для бесплатного сервиса получилась достаточно хорошая проверка, нашлось 4 ошибки, но только орфографические. Ресурс мультиязычный. При нахождении ошибки слово выделяется пунктирной линией, а при наведении на него курсором всплывают варианты замены, это удобнее, чем пошаговая проверка без возможности редактирования (рисунок 3):

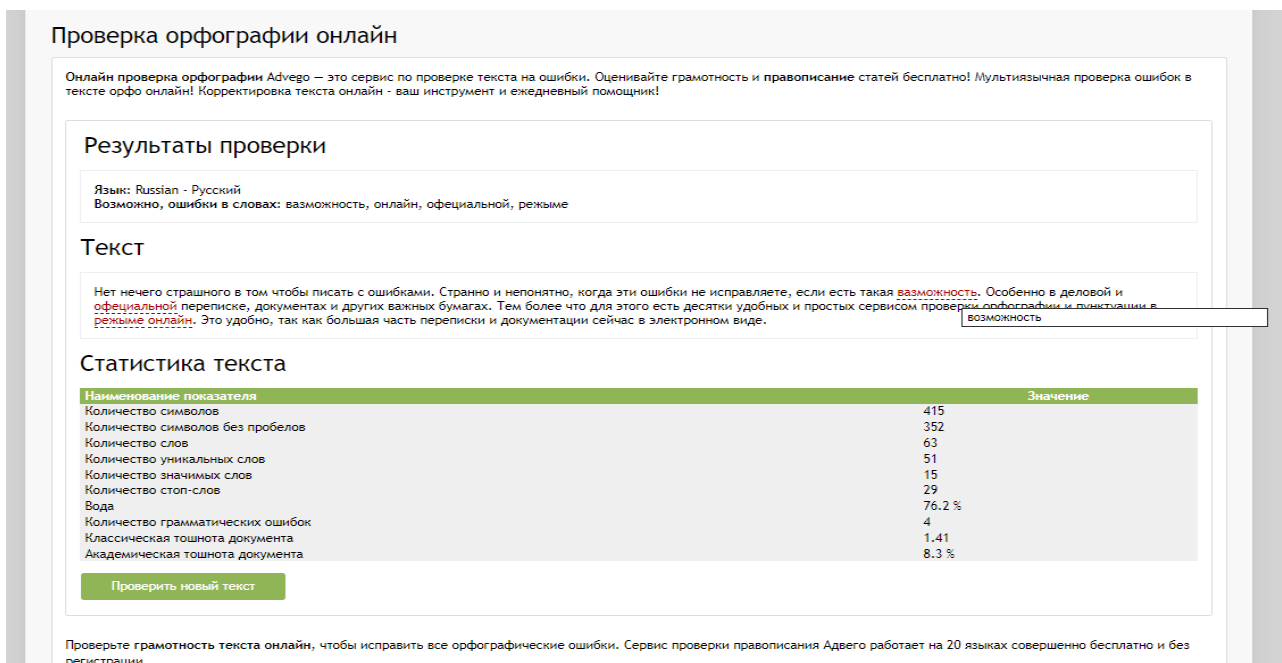


Рисунок 3 – Интерфейс работы сервиса ADVEGO

По результатам нашего небольшого анализа работы нескольких сервисов проверки русской орфографии можно сказать, что они дают достаточно схожие результаты, которые недостаточно точные. Некоторые ошибки не были распознаны. Мы рекомендуем для более точного исправления дорабатывать текст вручную. Однако мы считаем, что такие программы в любом случае являются очень полезными

*59-я научная конференция аспирантов, магистрантов и студентов БГУИР, 2023*

инструментами. В нашем исследовании сравнивалась производительность нескольких онлайн-сервисов проверки русской орфографии и было обнаружено, что Text.ru более удобен для пользователя. Однако текст всё равно необходимо проверять вручную, особенно при поиске и исправлении предложений, содержащих специальную лексику.

**Список использованных источников:**

1. Цветков, А. А. Использование информационных технологий в лингвистике [Электронный ресурс] / А. А. Цветков, Н. В. Демина. – Режим доступа: [https://interactive-plus.ru/ru/article/467879/discussion\\_platform](https://interactive-plus.ru/ru/article/467879/discussion_platform). – Дата доступа: 15.10.2022.
2. Орфография [Электронный ресурс]. – Режим доступа: <https://ru.wikipedia.org/wiki/Орфография>. – Дата доступа: 15.10.2022.
3. Проверка правописания [Электронный ресурс]. – Режим доступа: <https://roi4cio.com/categories/category/proverka-pravopisanija/>. – Дата доступа: 15.10.2022.
4. Text.ru [Электронный ресурс]. – Режим доступа: <https://text.ru/>. – Дата доступа: 15.10.2022.
5. LanguageTool [Электронный ресурс]. – Режим доступа: <https://languagetool.org/ru/>. – Дата доступа: 15.10.2022.
6. ADVEGO [Электронный ресурс]. – Режим доступа: <https://advego.com/text/>. – Дата доступа: 15.10.2022.