

Приложение **Текстоматика**

© 2021-2024 В.К. Иванов

Для чего это приложение?

Приложение **Текстоматика** предназначено для интеллектуальной оценки структуры и содержания текстовых документов. Является инструментом текстовых коммуникаций.

Что такое "оценивать текст"?

Оценить текст можно по-разному. Например, установить, что:

- объем документа превышает допустимый;
- документ содержит большое количество ошибок;
- содержание документа не соответствует заголовку;
- текст труден для понимания;
- не содержит необходимые разделы или не оформлен в соответствии с требуемым стандартом;
- и так далее.

В приложении **Текстоматика** оценка текста производится по этим и другим показателям, но более конкретно. Здесь вычисляются количественные значения заданных показателей качества текста. Вот несколько примеров:

- текст документа будет понятен специалисту на 65%;
- текст документа совпадает по смыслу с текстом документа другого автора на 92%;
- грамотность текста документа 15%.

Набор показателей качества текста задается в так называемой "модели оценки".

Кому и зачем надо оценивать тексты?

Профессиональная экспертиза текстовых документов является важной составляющей деятельности квалифицированных специалистов во многих областях.

Преподаватель вуза

- Проверяет тексты рефератов, контрольных, курсовых, дипломных и других письменных работ своих студентов.
- Формирует и обосновывает итоговую оценку проверенной работы.
- Оценивает тексты своих учебных и методических материалов с целью их совершенствования.

Школьный учитель

- Проверяет тексты творческих самостоятельных и контрольных работ, сочинений, изложений своих учеников.
- Формирует и обосновывает итоговую оценку проверенной работы.
- Оценивает тексты своих учебных и методических материалов с целью их совершенствования.

Редактор

- Проверяет авторские тексты (книги, статьи, обзоры и т.п.), приводя их в соответствие с установленными требованиями.

Рецензент

- Проверяет тексты рецензируемых материалов.
- Оценивает их соответствие установленным формальным и содержательным требованиям.
- Формирует и обосновывает итоговую оценку отрецензированного материала в рецензии (отзыве).

Эксперт

- Выполняет содержательную проверку текстовых документов, подлежащих экспертизе (соответствие содержания документа цели его подготовки, логическая завершенность, новизна и/или востребованность описанных объектов, уникальность текста и т.п.)
- Выполняет формальную проверку текстовых документов, подлежащих экспертизе (корректность структуры, состав реквизитов, размер документа и т.п.)
- Формирует и обосновывает итоговую оценку экспертизы документов в своем заключении.

Инженер, конструктор, технолог

- Контролирует соответствие проектной и эксплуатационной документации требованиям действующих стандартов.
- Оценит подготовленные материалы научно-технических мероприятий (такие как презентации, обзоры, пояснительные записки).

Исследователь, автор научных работ

- Оценивает соответствие своих текстов (научных статей, отчетов, книг, эссе, докладов, обзоров и т.п.) установленным формальным и содержательным требованиям.

Технический писатель

- Проверяет соответствие подготавливаемой проектной и эксплуатационной документации требованиям действующих стандартов.
- Оценит подготовленные подготовку проектной и эксплуатационной документации.

Аспирант и магистрант

- Оценивает соответствие своих текстов (научных статей, разделов диссертации и диссертации в целом, докладов, обзоров и т.п.) установленным формальным и содержательным требованиям.
- Улучшает тексты своих работ.

Студент университета или колледжа

- Проверяет тексты своих рефератов, отчетов по лабораторным работам, контрольных, курсовых, дипломных и других письменных работ на соответствие требованиям.
- Улучшает тексты своих работ.

Ученик школы

- Проверяет тексты своих письменных работ на соответствие требованиям.
- Улучшает тексты своих работ.

Почему оценка интеллектуальная?

Существует достаточное количество показателей качества текстов, вычисление которых возможно по тому или иному формальному алгоритму. Например, так можно вычислить показатель *удобочитаемости* текста или *степени соответствия текста заданной структуре*. Но в отношении многих показателей их вычисление является нетривиальной задачей, то есть не имеющей решения с помощью известных простых алгоритмов. Например, *обстоятельность* (полнота, доскональность, основательность) или *уникальность* слабоструктурированных и

семантически сложных текстовых объектов. В этом случае используются модели и методы искусственного интеллекта (такие как нейронные сети) и машинное обучение. В приложении **Текстоматика** мы применяем аналогичные подходы.

Зачем нужна модель оценки?

Модель оценки документов нужна для того, чтобы определить:

- Что мы будем оценивать, то есть задать характеристики текста, интересующие нас в данном случае.
- Как мы будем оценивать эти характеристики, то есть выполнить настройку параметров алгоритмов оценки.
- Насколько важна та или иная характеристика относительно других, то есть задать веса оцениваемых характеристик текста.

Кроме этого, модель оценки позволяет многократно выполнять оценку документов с использованием заданного набора оцениваемых характеристик.

Что конкретно оценивается?

Все показатели оценки документов делятся на группы:

- *Оценка содержания.* Примеры: соответствие содержания документа его заголовку, похожесть текста документа на тексты других документов.
- *Оценка текста.* Примеры: удобочитаемость текста, смысловая нагрузка текста.
- *Оценка представления.* Примеры: внутренняя структура документа, размер документа.
- *Оценка соответствия стандартам.* Примеры: оформление библиографического списка, количество ссылок, число цитирований в тексте.

Полный набор показателей, доступных в настоящее время для включения в модель оценки, можно увидеть при создании модели.

Как происходит оценка?

Оценка документов происходит следующим образом:

1. Один или несколько документов загружаются на сервер. При этом для каждого пользователя организуется цикл обработки его документов (сеанс), полностью изолированный от сеансов других пользователей.
2. Текст документа преобразуется в вид, удобный для анализа: выделяются разделы документа и их заголовки, абзацы, предложения, отдельные слова.
3. Выполняется анализ документа по каждому показателю, заданному в модели оценки. Используются специализированные алгоритмы, включая лексический и статистический анализ текста, регулярные выражения, нейронные сети.
4. Формируется отчет о результатах оценки документа по каждому показателю и документа в целом. В отчете содержатся **рекомендации** по совершенствованию текста документа. Отчет сохраняется в базе данных и может быть распечатан.

Почему именно Текстоматика?

Приложение **Текстоматика** привлекательно по следующим причинам:

- Большое количество алгоритмов оценки текстовых документов собраны в одном приложении и управляются из него. Во многих случаях нет необходимости обращаться к другим продуктам.
- Приложение свободно для использования и не требует платы за использование ключевых функций.
- Приложение не требует установки на устройство пользователя. Функции приложения доступны через браузер там, где есть Интернет.
- Сайт приложения адаптирован к мобильным устройствам (смартфонам, планшетам). Их можно использовать точно так же, как и настольный компьютер или ноутбук.
- В приложении поддерживаются все основные форматы текстовых файлов.
- Осуществляется гибкая настройка числа проверок и их параметров. Пользователь может настроить приложение на проверку документов по различным комбинациям показателей оценивания.
- Возможна оценка нескольких документов за один раз. Полезно для анализа и оценивания групп однотипных документов.

Примеры результатов оценки

Так выглядит веб-форма для показа результатов оценки текстового документа:

[https://www.ivkconsulting.ru/static/files/result_web.pdf \(/static/files/result_web.pdf?#zoom=auto&scrollbar=0&toolbar=1&navpanes=1\)](https://www.ivkconsulting.ru/static/files/result_web.pdf (/static/files/result_web.pdf?#zoom=auto&scrollbar=0&toolbar=1&navpanes=1))

А здесь можно посмотреть отчет о результатах оценки, предназначенный для печати:

[https://www.ivkconsulting.ru/static/files/result_pdf.pdf \(/static/files/result_pdf.pdf?#zoom=auto&scrollbar=0&toolbar=1&navpanes=1\)](https://www.ivkconsulting.ru/static/files/result_pdf.pdf (/static/files/result_pdf.pdf?#zoom=auto&scrollbar=0&toolbar=1&navpanes=1))

Начать работу с Текстоматика

1. Загрузить (/home) один или несколько документов.
2. Выбрать модель оценки или создать новую.
3. Оценить по выбранной модели текст одного или нескольких документов.
4. Изучить и сохранить при необходимости результаты оценки.

Разработка

- Идея, проектирование, разработка, документирование, продвижение: Владимир Константинович Иванов (<https://www.ivkconsulting.ru/?p=5577>).
- Опубликованные материалы, имеющие отношение к этому проекту, см. здесь: <https://www.ivkconsulting.ru/?p=4030&tgid=243#tppubs> (<https://www.ivkconsulting.ru/?p=4030&tgid=243#tppubs>).
- Другие проекты см. на сайте Исследования и разработки В.К. Иванова (<https://www.ivkconsulting.ru>) в разделе "Выполненные проекты": <https://www.ivkconsulting.ru/?p=3945> (<https://www.ivkconsulting.ru/?p=3945>)