

# Оглавление

<b>Введение</b>	<b>3</b>
<b>1 Анализ предметной области</b>	<b>4</b>
1.1 Задача упрощения текстов . . . . .	4
1.2 Актуальность . . . . .	5
1.3 Данные . . . . .	6
<b>2 Классификация существующих решений (методов решения, алгоритмов)</b>	<b>8</b>
2.1 Существующие подходы . . . . .	8
2.2 Решения абстрактного подхода . . . . .	8
2.3 Генерация нового текста . . . . .	9
2.3.1 Синтаксическое упрощение . . . . .	10
2.3.2 Статистический машинный перевод . . . . .	12
2.3.3 Использование методов глубокого обучения . . . . .	13
2.4 Метрики . . . . .	16
2.4.1 SARI . . . . .	16
2.4.2 Индексы удобочитаемости . . . . .	17
<b>Выводы</b>	<b>18</b>
<b>Список использованной литературы</b>	<b>19</b>

# Введение

Пишется в самом конце работ, является по сути кратким пересказом.

# 1 Анализ предметной области

В данном разделе вводятся основные определения и описываются важность и актуальность задачи упрощения текстов.

## 1.1 Задача упрощения текстов

Существует различные формулировки задачи упрощения текста. Так, в статье [1] даются определения в двух смыслах:

- упрощение текста в узком смысле - это процесс уменьшения лингвистической сложности текста при сохранении исходной информации и смысла;
- в более широком смысле упрощение текста охватывает другие операции: смысловое изменение для упрощения как формы, так и содержания; краткое изложение текста для исключения второстепенной или избыточной информации.

В статье [2] упрощением предложений называют процесс, целью которого является получение более легкого для чтения и понимания предложения за счет уменьшения его лексической и синтаксической сложности.

При этом задача упрощения текстов относится к области NLP (Natural language processing, обработка текстов на естественном языке) и имеет много общего с другими задачами из этой сферы - машинным переводом, рефразированием и обобщением (резюмированием) текста[3].

Важно отметить отличие упрощения текстов от их обобщения, так как эти задачи зачастую путают. Отличие заключается в том, что во втором случае основное внимание уделяется сокращению длины и содержания исходных данных. И хотя обобщенные тексты, как правило, короче, это не всегда так, и обобщение может привести к увеличению длины полученных предложений[11]. В рамках же упрощения текста обычно сохраняется все содержание.

## 1.2 Актуальность

В последние десятилетия количество неструктурированных текстовых данных резко возросло в связи с появлением Интернета. Как следствие, возросла и потребность в их упрощении, что показано на рисунке 1.1.

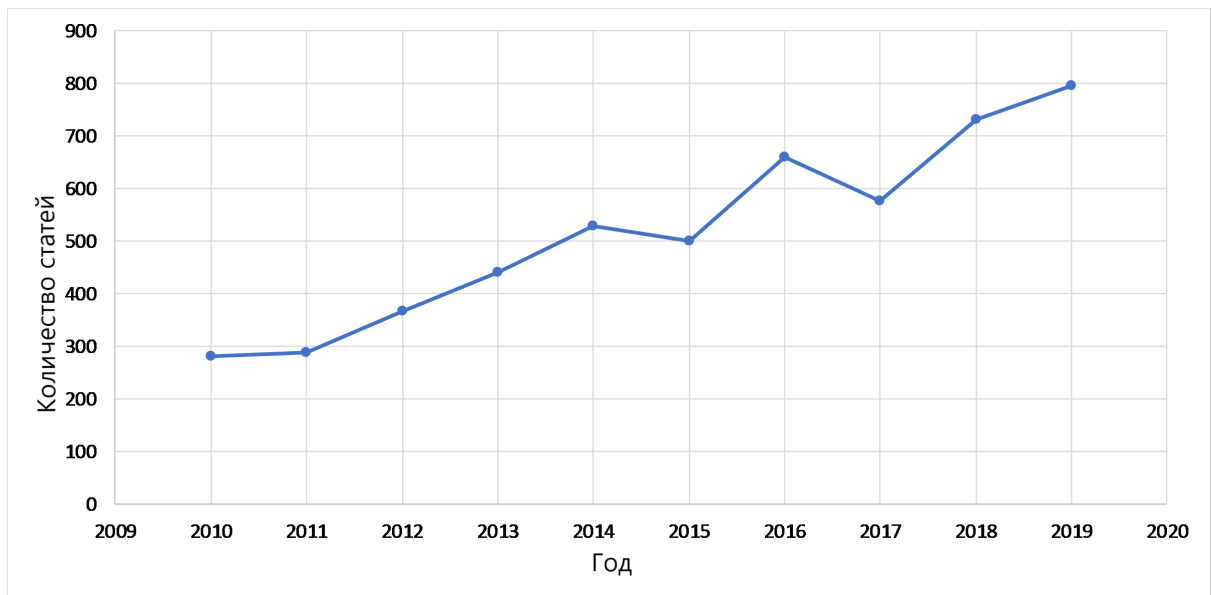


Рисунок 1.1 – График роста интереса к теме упрощения текста. Создано на основе статистики Google Scholar по поисковым запросам "Text Simplification"("Упрощение текстов"), "Lexical Simplification"("Лексическое упрощение"), "Syntactic Simplification"("Синтаксическое упрощение") [4].

Упрощение текстов необходимо для различных задач и целевых аудиторий:

- в качестве этапа подготовки текста перед его обобщением [5];
- для людей, изучающих иностранный язык, и детей, учащихся читать (требуется лексическое упрощение для сокращения количества специализированных и нечастотных слов) [6];
- для людей с дислексией и афазией, для которых особенно длинные слова и предложения могут представлять трудности;
- для людей, страдающих аутизмом (необходимо уменьшать количество образных выражений и синтаксическую сложность) [7].

## 1.3 Данные

Упрощение предложений привлекло многих исследователей в связи с появлением больших параллельных корпусов. Наиболее известные из них

составлены из английских текстов, например, PWKP и Wiki-large. Последний представляет собой большой параллельный корпус на английском языке, состоящий из сложных предложений, взятых из Википедии, и их выровненных упрощенных версий.

Подобных данных на других языках значительно меньше. Большой корпус на русском языке был собран, когда задача упрощения текстов была предложена в рамках международной конференции по компьютерной лингвистике и интеллектуальным технологиям DIALOGUE 2021<sup>1</sup>. Организаторы подготовили обучающие и тестовые наборы на русском языке с использованием краудсорсинговой платформы, а также перевели тексты из Википедии (корпус RuWikiSimple). Другим источником данных могут стать результаты перевода в виде перефразирования[8].

Основная проблема в упомянутых данных заключается в том, что имеет место фокусировка на текстах Википедии, что ограничивает исследования и приводит к неадекватности моделей на других типах данных [8].

## Выводы из анализа предметной области

Таким образом, актуальность задачи упрощения текстов в последнее десятилетие увеличивается, формируются новые корпуса на различных языках для обучения моделей. При этом поставленная задача схожа с другими задачами из сферы NLP и если использовать ее более широкое понятие, то она будет трудно отличима от задачи обобщения. Поэтому, чтобы разграничить эти два понятия, в данной работе будет использоваться более узкое определение задачи упрощения.

---

<sup>1</sup><http://www.dialog-21.ru/dialogue2021/results/>

## 2 Классификация существующих решений (методов решения, алгоритмов)

Описываются существующие решения, даются ссылки. Предлагаются критерии оценки методов. Хорошо, если критерии имеют обоснование. Приводится классификация решений по критериям в виде таблицы, но в отдельных случаях можно и не в виде таблицы (пример мы Вам кидаем отдельно от документа).

В данном разделе описываются существующие решения задачи упрощения текстов, предлагаются критерии оценки методов и приводится классификация решений по этим критериям.

### 2.1 Существующие подходы

Глобально выделяют два подхода к решению задачи упрощения текстов - экстрактивный (извлекающий) и абстрактный.

Большинство ранних работ, посвященных задаче упрощения текстов, использовали экстрактивный подход - выделение в документе тех предложений, которые передают больше информации. Этот подход достаточно прост в реализации, однако он подходит лишь для решения задачи упрощения текстов в широком смысле, поэтому в дальнейшем в данной работе рассматриваться не будет.

С ростом доступности вычислительных ресурсов и количества исследований в области NLP для решения задачи упрощения текстов все чаще стал использоваться абстрактный подход, подразумевающий генерацию нового текста [9].

### 2.2 Решения абстрактного подхода

Решения задачи упрощения текстов, относящиеся к абстрактному подходу, можно разделить на решения в виде лексического упрощения и в виде генерации нового текста.

Первоначально абстрактный подход подразумевал упрощение предложений за счет лексических замен на уровне слов или целых фраз [10]. Этот процесс фокусируется исключительно на сокращении лексического содержания текста, но не принимает во внимание такие подзадачи, как грамматическое или синтаксическое упрощение [11]. Таким образом, эти решения не учитывают все аспекты решаемой задачи и поэтому в дальнейшем в работе рассматриваться не будут.

Современные решения, основанные на абстрактном подходе, включают в себя разбиение сложных предложений на более простые, удаление редко употребляемых слов и генерацию на этой основе нового текста. Это стало возможным благодаря появлению нейронных сетей, в частности, рекуррентных нейронных сетей (RNNs)<sup>1</sup>, которые позволяют решать задачи «от последовательности к последовательности» (seq2seq)<sup>2</sup>.

Последовательность действий в решении задач, сведенных к seq2seq, является универсальной. Предварительная обработка данных включает очистку входного и результирующего текста, удаление знаков препинания и специальных символов, формирование словаря типичных слов. Далее входные и результирующие предложения преобразовываются в числовую форму с вектором одинаковой длины либо путем их усечения, либо путем их дополнения. Затем модель обучается, после чего новые поступающие данные также сначала векторизуются, преобразуются к определенной длине и упрощаются, а только после этого преобразовываются обратно из числовой формы в текстовую.

## 2.3 Генерация нового текста

Далее будут рассмотрены различные виды решений задачи упрощения текста, относящиеся к классу тех, которые используют генерацию нового текста.

---

<sup>1</sup>Рекуррентная нейронная сеть, или RNN, - это сеть, которая работает с последовательностью и использует собственные промежуточные выходные данные в качестве входных данных для последующих шагов [12]

<sup>2</sup>Сеть sequence-to-sequence («от последовательности к последовательности»), или сеть seq2seq, или сеть кодировщика-декодера, представляет собой модель, состоящую из двух RNN, называемых кодировщиком и декодером. Кодер считывает входную последовательность и выдает один вектор, а декодер считывает этот вектор для создания выходной последовательности [12]



### 2.3.1 Синтаксическое упрощение

Цель синтаксического упрощения заключается в выявлении грамматически сложных частей текста и их переписывание для облегчения понимания. Такое упрощение может включать разделение длинных предложений на более короткие фрагменты, переписывание предложений со страдательным залогом так, чтобы в них использовался залог действительный<sup>3</sup>, разрешение двусмысленностей и анафор [11]. По ходу выполнения синтаксического упрощения удастся заменять слова, которые являются сложными из-за наложения морфем, участвующих в переводе слова из одной части речи в другую, на их более простые, оригинальные версии.

Основополагающей работой в области синтаксического упрощения была система автоматического создания правил переписывания предложений [14], которая брала аннотированные корпуса и изучала возможные принципы упрощения для конкретной предметной области.

Более поздние работы по синтаксическому упрощению были сосредоточены на улучшении структуры выходного текста - обеспечении того, чтобы предложения появлялись в правильном порядке [15]. Также этот подход стали использовать для распознавания именованных сущностей (Named Entity Recognition, NER), особенно в области медицины [16].

Синтаксическое упрощение обычно выполняется в три этапа. Пример применения каждого этапа пример показан на рисунке 2.1

1. Сначала текст анализируется для определения его структуры и создания дерева синтаксического анализа. Это может быть сделано на разных уровнях детализации, но наилучшие результаты достигаются на довольно грубом уровне, когда слова и фразы группируются в так называемые «супер-теги», представляющие собой фрагменты исходного предложения. Такие теги могут быть объединены по обычным

---

<sup>3</sup> Действительный залог имеют глаголы переходные, обозначающие действие, производимое субъектом и активно направленное на объект. Действительный залог имеет синтаксическую характеристику: субъект действия является подлежащим, а объект - дополнением в винительном падеже без предлога: Мир победит войну.

Страдательный залог выражается присоединением к глаголам действительного залога аффикса -ся (ср.: Рабочие строят дома. - Дома строятся рабочими). Кроме того, значение страдательного залога может быть выражено формами страдательных причастий - полных и кратких. Например: Мать любима (любимая). Тема изучена (изученная)[13].

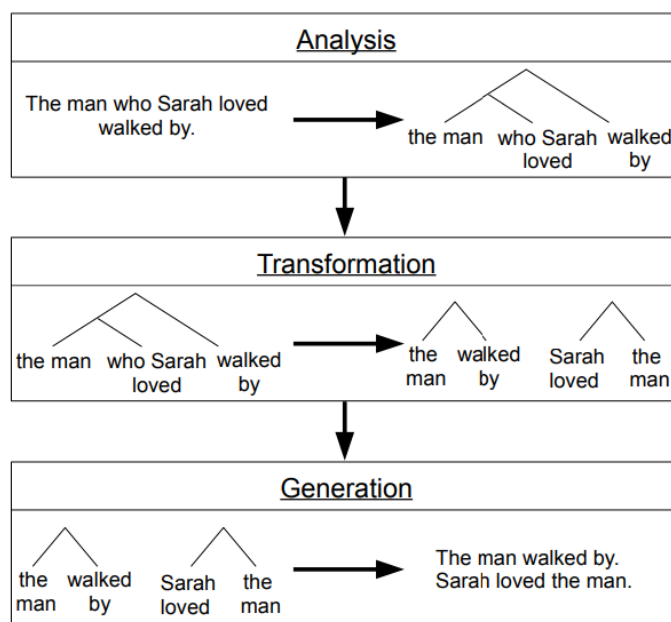


Fig. 3. The syntactic simplification pipeline, with worked example. Pre-determined rewrite rules govern the simplifications that occur during the transformation step. The generation step is important to ensure the cohesion of the resultant text.

Рисунок 2.1 – todo

грамматическими правилами и являются структурированной версией текста. На этом же этапе простой проверкой по заранее заданным правилам или с помощью бинарного классификатора SVM<sup>4</sup> [11] определяется сложность предложения.

2. На этапе преобразования дерево синтаксического анализа модифицируется в соответствии с набором правил, которые выполняют операции упрощения. Например, разбиение сложного предложения на более простые, перестановка или удаление этих простых предложений[18].
3. Далее следует фаза восстановления предложения, в ходе которой в текст вносятся дополнительные изменения для улучшения согласованности и удобства чтения.

Синтаксическое упрощение считалось важным компонентом систем упрощения текстов и было реализовано в системах, которые повсеместно ис-

<sup>4</sup> «Машина опорных векторов» (Support Vector Machine, SVM) - это алгоритм машинного обучения с учителем, который в основном используется в задачах классификации. Каждая запись представляется в виде точки в n-мерном пространстве (где n - количество признаков), при этом значение каждого признака равно значению определенной координаты. Затем выполняется классификация, в результате которой находится гиперплоскость, разделяющая два класса [17].

пользуются как вспомогательные, например, в PSET[?] и Porsimples[19].

Преимущества синтаксического упрощения заключаются в его высокой точности и применимости к другим задачам NLP [11]. Недостатком является трудоемкость создания и проверки применимости правил перезаписи. В последнее время достижения в области методов глубокого обучения привели к автоматизации процесса обнаружения возможности применения синтаксического упрощения.

## 2.3.2 Статистический машинный перевод

Автоматизированный машинный перевод является устоявшейся техникой в NLP. Эта задача подразумевает автоматическое преобразование лексики и синтаксиса одного языка в синтаксис другого, в результате чего получается переведенный текст. Машинный перевод был успешно применен[11] к задаче упрощения текстов путем ее переформулирования в задачу «одноязычного перевода». То есть задача упрощения сводится к переводу с исходного «сложного» русского языка на целевой «простой» русский язык.

Разновидностью машинного перевода является статистический машинный перевод (Statistical Machine Translation (SMT)), основанный на статистических моделях, которые изначально «ничего не знают о правилах и лингвистике», а затем изучают большие объемы пар предложений из выровненных двуязычных корпусов, настраивая свои параметры (наиболее вероятный вариант перевода того или иного слова), а затем применяются к новым текстам.

Например, модель для перевода с русского на английский изучила перевод одного предложения: «Я вижу дом» в «I see a house». Если теперь запросить у нее перевод слова «дом», она предположит, что слово с равной вероятностью переводится как «I», «see», «a» или «house». Но если предоставить модели еще одно соответствие, что предложение «Этот дом большой» переводится как «That house is big», то из анализа уже двух сопоставлений переводчик отметит, что в переводах все слова встретились по одному разу, а «house» — дважды, равно как и слово «дом» (и никакое другое) в исходных предложениях. А значит, по сравнению со всеми остальными вариантами увеличивается вероятность соответствия «дом =

house» и между ними установилась связь.

Данная задача облегчается, когда исходный и целевой языки схожи, и для перевода предложений требуется минимальное число изменений его структуры. И именно этот тип машинного перевода был применен к задаче упрощения текстов[11].

Эффективность применения статического машинного перевода для задачи упрощения предложений в значительной степени зависит от набора данных, используемых для обучения модели. Например, если в них содержатся слишком длинные исходные предложения или упрощенные предложения слишком сильно отличаются по своей структуре от входных, то этот подход не позволит отслеживать и верно сопоставлять различные части предложений. Кроме того, данный метод не учитывает знаки препинания и разбиение сложных предложений, что приводит к потере основного контекста.

### **2.3.3 Использование методов глубокого обучения**

Глубокое обучение - это разновидность машинного обучения, в которой нейронные сети и алгоритмы, основанные на структуре и функционировании человеческого мозга, обучаются на большом объеме данных для создания шаблонов принятия решений. Этот подход позволяет обучаться путем многократного выполнения задач и настройки модели для улучшения результата[20].

Методы глубокого обучения после своего появления стали активно применяться для решения задачи упрощения текстов, сформулированной в терминах моделирования seq2seq. Однако в ранних моделях seq2seq были две существенные проблемы.

Во-первых, это неточность результата. Эффективность моделей кодировщика-декодера сильно зависит от расположения слов в исходном предложении, поэтому модель зачастую не может расположить редко употребляемые слова в корректную позицию выходного предложения. Один из вариантов решения этой проблемы - добавление «указателя» на подобные слова в исходном тексте[21]. Такое решение показало многообещающие результаты в сохранении точного контекста сгенерированного упрощенного текста.

Во-вторых, это повторения в выводе. Данная проблема часто возникает в простых моделях seq2seq из-за того, что так называемые «стоп-слова» («как», «и», «а», «то» и т. д.) встречаются в тексте намного чаще остальных, и модель учится чаще предсказывать эти слова. В частности, именно эта проблема стала основным недостатком в решении, описанном ранее[21]. Для борьбы с этим недостатком было предложено штрафовать модель за повторения с помощью введение векторов «покрытия», «внимания» и «контекста». Эти вектора отслеживают слова, которые передаются из исходного предложения в упрощенное: вектор «внимания» - слова, несущие основной смысл, «контекста» - сопутствующую информацию, «покрытия» - общий переданный объем слов. Именно вектор покрытия призван дополнительно контролировать вектора «внимания» и «контекста» и штрафовать модель при их наложении друг на друга[9].

Также к задаче упрощения текстов была успешно применена техника обучения с подкреплением[22]. Была разработана модель кодировщика-декодера в сочетании с системой глубокого обучения с подкреплением DRESS (сокр. от Deep Reinforcement Sentence Simplification, глубокое упрощение предложений с подкреплением), стремящаяся оптимизировать функцию потерь, которая поощряет простые, легко читаемые и сохраняющие исходный смысл результаты упрощения. С помощью этой модели было показано, что обучение с подкреплением предоставляет возможность для предоставления дополнительной (предварительной) информации в данные.

Все топовые модели использовали ту или иную форму фильтрации обучающего набора данных или обработки управляющих маркеров и тонкой настройки крупномасштабных предварительно обученных языковых моделей. Выигрышное решение (qbic) в значительной степени основано на упрощении многоязычных предложений без пристрастия [16]. Модель состоит из MBART[19], точно настроенной на Paraphraserplus[9] и RuWikiSimple, основанной на определенных контрольных маркерах (сходство Левенштейна, доля совпадающих символов между оригинальными и упрощенными предложениями, ранг слова, сходство лексем). Несколько других моделей, занявших первое место (оржан, занявший второе место, ашатилов, занявший третье место, и аленуш, занявший пятое место), являются генеративными (модели на основе GPT), точно настроенными на отфильтрованном

примере RuWikiSimple. Чтобы быть более конкретным, второй-помещается модель orzhan является ruGPT-3 доработаны на RSSE Дев набора и фильтруют RuWikiSimple, где фильтрация проводилась с помощью 6 различных метрик (предложении встраивание Косинус сходство, именованных сущностей результат сохранения, лексических результат сложности, зависимость глубина дерева счет, счет длины и чтения). Выбор лучшего кандидата из предложенных вариантов сгенерированный моделью путем оптимизации комбинации этих шести показателей вместо SARI, учитываемых для Улучшение на 0,6 SARI в частном тесте.

Решение Ашатилова, занявшее третье место, использовало модель на основе GPT-2[13]. Как простая фильтрация Ruwik, так и выбор кандидатов выполнялись с помощью четырех показателей (косинусное сходство, ROUGE-L и длина ввода и кандидата в токенах). Однако, в отличие от модели оржана, вместо ручного объединения четырех показателей в агрегат ашатилов выбирает лучшего кандидата путем обучения классификатора случайных лесов, где в качестве признаков используются четыре показателя.

Несколько других решений на основе mBART обогащены некоторыми дополнительными функциями и технологиями. Они включали предварительную подготовку по дополнительным источникам данных (наиболее популярным из которых был ParaPhraserPlus), различные функции ручной работы и обратный перевод[26].

Анализируя результаты, можно предположить, что выбор между предварительной подготовкой seq2seq (т. е. на основе MBART) и генеративными моделями (на основе GPT) оказывает ограниченное влияние на конечный результат. Использование дополнительных показателей для фильтрации набора данных, отбора кандидатов и/или в качестве контрольных маркеров, наоборот, представляется крайне важным для дальнейшего повышения производительности. Похоже, что выбор показателей в двух лучших моделях лучше, чем те, которые используются в третьей модели. С другой стороны, обучение отдельной модели для выбора лучшего кандидата (как это делается моделью, занявшей третье место), по-видимому, приводит к меньшему переобучению, чем использование совокупной метрики с фиксированными параметрами (о чем свидетельствует сокращение моде-

ли ашатилова на 0,4 САРИ по сравнению с моделью Ашатилова до 1 САРИ оржана и аленуша). Для подтверждения этих утверждений необходимо провести дополнительные исследования.

## 2.4 Метрики

Оценка качества упрощения текста - довольно субъективная задача, трудно переводимая на язык компьютера. Поэтому до сих пор предпочтение отдается человеку, который оценивает упрощенное предложение с точки зрения его грамматической корректности, сохранения смысла и простоты, используя шкалы Лайкерта<sup>5</sup>. Однако тенденция к использованию технологии обучения без учителя потребовала разработки или адаптации уже существующих показателей для автоматической оценки качества упрощения.

### 2.4.1 SARI

Метрика SARI (system output **a**gainst **r**eferences and against the **i**nput sentence - результат системы против эталонного решения и исходного предложения) была разработана в 2016 году специально для задачи упрощения и в настоящее время считается наиболее применимой[24].

Авторы SARI заметили, что в отличие от метрик качества машинного перевода, где исходное предложение записано на другом языке, метрики качества упрощения могут использовать еще один источник для оценки результирующего предложения - само исходное предложение, так как оно записано на том же языке, что и упрощенное. Разработанная ими метрика явно оценивает, насколько корректно были выбраны слова, которые были добавлены, удалены или же сохранены моделью.

Результаты исследования авторов показали высокую корреляцию оценки метрикой SARI с оценками качества упрощения человеком, однако, эта

---

<sup>5</sup>Шкала Лайкерта - это ординальная (порядковая) шкала ответов на вопрос или утверждений, расположенных в иерархической последовательности, например, от «полностью согласен» через «затрудняюсь ответить» и до «категорически не согласен»[23].

функция требует нескольких «справочных» примеров упрощения, которые не всегда доступны.

## 2.4.2 Индексы удобочитаемости

Индексы удобочитаемости используются в США для присвоения любому тексту уровня его «простоты». Обычно для такой оценки применяется сразу несколько формул, которые учитывают количество предложений, слогов, общее количество слов и количество редких слов в рассматриваемом тексте. Отличаются эти формулы лишь коэффициентами, расположением членов в формуле или способом интерпретации результата.

Например, в тесте Флэша-Кинкайда предполагается, что чем меньше слов в предложениях и чем короче слова, тем более простым является текст. В результате применения формулы 2.1 корпус получает оценку, по которой с помощью специальной таблицы интерпретируется уровень образования, необходимый для понимания текста[25].

$$206.835 - 1.015 \cdot \left( \frac{\text{количество слов}}{\text{количество предложений}} \right) - 84.6 \cdot \left( \frac{\text{количество слогов}}{\text{количество слов}} \right) \quad (2.1)$$

Основной недостаток индексов удобочитаемости заключается в том, что используемые в них формулы «поощряют» короткие предложения с простыми словами, но не способны распознавать грамматически неправильные результаты или упрощенные предложения, некорректно передающие смысл исходного.

– SAMSA - A new metric proposed by Sulem et al. that can evaluate simplification, including sentence splitting, not just summarization, like SARI [89]. Due to its recent release, this metric is not currently widely used.



# Выводы

Как ни странно, описывается не работа уже в третий (потому что второй раз она была описана во введении) раз, а даётся формальный ответ на вопрос, было ли выполнено ТЗ: были выработаны и обоснованы критерии оценки, был предложен метод решения...

# Литература

- [1] Siddharthan Advaith. A survey of research on text simplification.
- [2] MUSS: Multilingual Unsupervised Sentence Simplification by Mining Paraphrases / Louis Martin, Angela Fan, Éric de la Clergerie [и др.]. URL: <http://arxiv.org/abs/2005.00352>.
- [3] Zhu Zhemin, Bernhard Delphine, Gurevych Iryna. A Monolingual Tree-based Translation Model for Sentence Simplification. C. 1353–1361. URL: <https://aclanthology.org/C10-1152>.
- [4] Sikka Punardeep, Mago Vijay. A Survey on Text Simplification. URL: <http://arxiv.org/abs/2008.08612>.
- [5] Finegan-Dollak Catherine, Radev Dragomir R. Sentence simplification, compression, and disaggregation for summarization of sophisticated documents. T. 67, № 10. C. 2437–2453. URL: <https://onlinelibrary.wiley.com/doi/10.1002/asi.23576>.
- [6] Liu Jun, Matsumoto Yuji. Simplification of Example Sentences for Learners of Japanese Functional Expressions // Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA2016). The COLING 2016 Organizing Committee. C. 1–5. URL: <https://aclanthology.org/W16-4901>.
- [7] Evans Richard, Orăsan Constantin, Dornescu Iustin. An evaluation of syntactic simplification rules for people with autism. C. 131–140. URL: <https://aclanthology.org/W14-1215>.
- [8] RuSimpleSentEval-2021 Shared Task: Evaluating Sentence Simplification for Russian / Kazan Federal University, Kazan, Russia, Andrey Sakhovskiy, Alexandra Izhevskaya [и др.]. C. 607–617. URL: <http://www.dialog-21.ru/media/5558/sakhovskiyaplusetal161.pdf>.
- [9] See Abigail, Liu Peter J., Manning Christopher D. Get To The Point: Summarization with Pointer-Generator Networks // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics

- (Volume 1: Long Papers). Association for Computational Linguistics. C. 1073–1083. URL: <https://aclanthology.org/P17-1099>.
- [10] A Survey on Lexical Simplification. T. 60. URL: <https://www.jair.org/index.php/jair/article/view/11091>.
  - [11] Shardlow Matthew. A Survey of Automated Text Simplification. T. 4.
  - [12] NLP From Scratch: Translation with a Sequence to Sequence Network and Attention — PyTorch Tutorials 1.10.1+cu102 documentation. URL: [https://pytorch.org/tutorials/intermediate/seq2seq\\_translation\\_tutorial.html](https://pytorch.org/tutorials/intermediate/seq2seq_translation_tutorial.html).
  - [13] Валгина Н.С., Розенталь Д.Э., Фомина М.И. Современный русский язык: Учебник / Под редакцией Н.С. Вал. URL: [https://pedlib.ru/Books/6/0262/6\\_0262-288.shtml#book\\_page\\_top](https://pedlib.ru/Books/6/0262/6_0262-288.shtml#book_page_top).
  - [14] Chandrasekar R., Srinivas B. Automatic Induction of Rules for Text Simplification.
  - [15] Siddharthan Advaith. Syntactic Simplification and Text Cohesion. T. 4, № 1. C. 77–109. Number: 1 Publisher: Springer Netherlands. URL: <http://oro.open.ac.uk/58888/>.
  - [16] Jonnalagadda Siddhartha, Gonzalez Graciela. BioSimplify: an open source sentence simplification engine to improve recall in automatic biomedical information extraction. T. 2010. C. 351–355. URL: <http://www.scopus.com/inward/record.url?scp=84964959827&partnerID=8YFL>.
  - [17] SVM | Support Vector Machine Algorithm in Machine Learning. URL: <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/>.
  - [18] Barlacchi Gianni, Tonelli Sara. ERNESTA: A Sentence Simplification Tool for Children’s Stories in Italian // Computational Linguistics and Intelligent Text Processing / под ред. Alexander Gelbukh. Springer Berlin Heidelberg. T. 7817. C. 476–487. Series Title: Lecture Notes in Computer Science. URL: [http://link.springer.com/10.1007/978-3-642-37256-8\\_39](http://link.springer.com/10.1007/978-3-642-37256-8_39).

- [19] Aluísio Sandra, Gasperin Caroline. Fostering Digital Inclusion and Accessibility: The PorSimples project for Simplification of Portuguese Texts // Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas. Association for Computational Linguistics. C. 46–53. URL: <https://aclanthology.org/W10-1607>.
- [20] A Review on Deep Learning Techniques Applied to Semantic Segmentation / Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea [и др.]. URL: <http://arxiv.org/abs/1704.06857>.
- [21] Exploring Neural Text Simplification Models / Sergiu Nisioi, Sanja Štajner, Simone Paolo Ponzetto [и др.]. C. 85–91. URL: <https://aclanthology.org/P17-2014>.
- [22] Zhang Xingxing, Lapata Mirella. Sentence Simplification with Deep Reinforcement Learning. C. 584–594. URL: <https://aclanthology.org/D17-1062>.
- [23] отв. ред. Г. В. Осипов Л. Н. Москвичев; уч. секр. О. Е. Чернощек. Косолапов М.С. Шкала Лайкерта (Ликерта). Социологический словарь // М.: Академический учебно-научный центр РАН-МГУ им. М.В. Ломоносова, НОРМА, НИЦ ИНФРА М. 2015. С. 372–373.
- [24] Optimizing Statistical Machine Translation for Text Simplification / Wei Xu, Courtney Napoles, Ellie Pavlick [и др.]. Т. 4. С. 401–415. URL: [https://doi.org/10.1162/tacl\\_a\\_00107](https://doi.org/10.1162/tacl_a_00107).
- [25] Paasche-Orlow Michael K., Taylor Holly A., Brancati Frederick L. Readability Standards for Informed-Consent Forms as Compared with Actual Readability. Т. 348, № 8. С. 721–726. Publisher: Massachusetts Medical Society \_eprint: <https://doi.org/10.1056/NEJMsa021212>. URL: <https://doi.org/10.1056/NEJMsa021212>.