

## **Техническое задание Проект “Автотьютор”**

Научные руководители: О.С. Еремина, А.А. Бонч-Осмоловская

Команда: магистры программы “Компьютерная лингвистика” Н. Бабаков, Э. Гонг, Н. Исупова, Е. Цветков, Ц. Ма; магистры программы “Русский язык как иностранный” Д. Илюхина, О. Чекмачева

### **ЦЕЛЬ И ЗАДАЧИ**

Цель: подбирать аутентичные русскоязычные тексты для чтения, которые будут соответствовать уровню студента, который изучает русский язык как иностранный.

Задачи:

- исследовательские:
  - справедливо ли предположение о необходимости отказаться от уровневой системы анализа текстов в контексте рекомендательной системы текстов для изучающих русский как иностранный
  - какие текстовые фичи имеют наибольшее влияние на уровень понимания учеником текста, как такие фичи автоматизировано извлекать из текста
  - как анализировать лексический запас учеников, как сортировать лексику, какие методы применять
- инженерные - что за продукт мы хотим создать
  - как извлечь текстовые фичи, которые будут признаны самыми эффективными
  - как эффективно реализовать тестирование/я с программной точки зрения
  - как выстроить процесс подбора текста, откуда брать сами тексты

### **ДАННЫЕ И МЕТОДЫ**

Диссертации и статьи о readability

- “Отбор релевантных признаков для автоматического определения сложности русского текста как иностранного” Лапошина А.Н.
- Russian natural language processing for computer-assisted language learning. Robert J Reynolds
- A Readable Read: Automatic Assessment of Language Learning Materials based on Linguistic Complexity. Pitan, Vajjala, Volodina
- Assessing the readability of fiction: a corpus analysis and readability ranking of 200 English fiction texts. Dalvean, Enkhbayar
- Automated Text Readability Assessment for Russian Second Language Learners. Laposhina, Veselovskaya, Lebedeva, Kupreshchenko
- Self-assigned ranking of L2 vocabulary: Using the Bricklayer computer game to assess depth of word knowledge. Brumbaugh
- Characterizing Text Difficulty with Word Frequencies. Chen, Meuers

- Automatic Readability Assessment. Lijun Feng
- Matching Readers to Texts: An Approach that Maintains Construct Representation. Sheehan
- Online Readability and Text Complexity Analysis with TextEvaluator. Diane Napolitano, Kathleen M. Sheehan, and Robert Mundkowsky
- Russian natural language processing for computer-assisted language learning, Russian natural language processing for computer-assisted language learning. Robert Reynolds
- Single-sentence Readability Prediction in Russian. Nikolay Karpov, Julia Baranova, Fedor Vitugin
- The TextEvaluator Tool: Helping Teachers and Test Developers Select Texts for Use in Instruction and Assessment. Kathleen M. Sheehan,
- Validating Automated Measures of Text Complexity. Kathleen M. Sheehan
- On Improving the Accuracy of Readability Classification using Insights from Second Language Acquisition. Sowmya Vajjala, Detmar Meurers
- Readability Classification for German using lexical, syntactic, and morphological features. Julia Hancke, Sowmya Vajjala, Detmar Meurers,
- Помощь от проекта Easy Ten

#### Документы ТРКИ:

- Лексический минимум (уровни ТРКИ-1, ТРКИ-2, ТРКИ-3)
- Государственные образовательные стандарты по русскому как иностранному (уровни ТРКИ-1, ТРКИ-2, ТРКИ-3)

#### Тексты для последующей рекомендации

- Национальный Корпус Русского Языка
- Пособия по русскому языку, подготовленные для определенного уровня учеников
- Примеры тестов РКИ для составления аналогичных тестов на выявлениях грамматических или лексических знаний ученика

#### Методы:

- Полевое исследование посредством проведения теста, направленное на 1) выявление частоты ошибок на тех или иных словах / грамматических конструкциях 2) определение уровня учащегося по нескольким шкалам (знание лексики, грамматики и др.) для дальнейшей рекомендации
- Выделение фич текста с помощью Python, алгоритмов и уже реализованных библиотек/парсеров

## ЭТАПЫ

1. Исследование параметров, влияющих на сложность текста - ноябрь 2018 - январь 2019
  - Грамматика
    - КЛ. Поиск наиболее часто применяемых в аналогичных исследованиях текстовых фич
    - РКИ. Составление таблицы текстовых фич по стандартам ТРКИ

- Приведение фич к общему знаменателю.
- Лексика
  - КЛ. Исследование методов тестирования лексики. Поиск лучшего способа ранжирования лексики.
  - РКИ. Предложение альтернативных подходов к анализу восприятия слова. К примеру, уровень абстракции и так далее.

2. Реализация теста на определение уровня ученика, сбор дополнительных данных, построение модели рекомендации текстов с помощью полученных параметров, тестирование модели - январь - апрель 2019

- КЛ. Тест на выявление сложности лексики. В зависимости от типа теста реализация программной платформы для его проведения, при отсутствии необходимости - простая подготовка структуры теста
- РКИ. Контентное наполнение теста и его непосредственное проведение.
- КЛ. Анализ данных по сложности лексики. Составление и программная реализация финального теста (грамматика и лексика) для определения уровня знаний ученика. Разметка текстов для дальнейшей рекомендации на основании прохождения теста и изученных параметров
- РКИ. Проведение теста на выявление уровня знаний.
- КЛ. Построение модели рекомендации текстов на основе пройденного теста с учетом выявленных параметров.
- РКИ. Экспертная оценка соответствия сложности подобранных текстов уровню знаний ученика

3. Реализация прототипа системы, написание исследовательской статьи - апрель - июнь 2019

- КЛ. По возможности - реализация прототипа системы тестирования и подбора текста с минимальным интерфейсом.
- КЛ и РКИ. Написание исследовательской статьи по итогам проекта.

	Этап	Бабаков	Исупова	Цветков	Гонг	Илюхин а	Чекмачева
январь	пар-ры текста	лексика	граммат ика	граммат ика	лексика	таблица ТРКИ	сложность лексики
февраль апрель	тест сложности лексики	логика теста	-	-	логика теста	контент, организа ция тестиро вания	организац ия тестирова ния
февраль апрель	тест уровня ученика	платфор ма для теста	платфор ма для теста	разметк а текстов	платфор ма для теста	организа ция тестиро	контент, ор ганизация тестирова

				для рекомен дации		вания	ния
февраль апрель	Оценка результато в	валидац ия рез-тов	валидац ия рез-тов	валидац ия рез-тов	валидац ия рез-тов	индивидуал ьный разбор текстов с выборкой учеников	индивидуальн ый разбор текстов с выборкой учеников
май - июнь	Готовый продукт	==	==	==	==		
	Статья	==	==	==	==	==	==