Результаты совещания от 20.05.20

Было принято решение, на ранних стадиях разработать интерфейс, через который возможно взаимодействие с ИС в отдельном приложении.   
В дальнейшем, когда будет достигнут нужный уровень обучения нейронной сети и будет разработан соответствующий функционал, рассматривается перенос на “Образовательный портал“.

Интерфейс должен включать в себя:

1. Раздел для загрузки файла статьи
2. Поля для вывода и корректировки (при необходимости) ключевых слов, предложенных ИС
3. Информационные поля, в которых после загрузки статьи должно быть отображено:
   1. К какой тематике относится загруженная статья
   2. Что можно было бы добавить/исправить (В перспективе)

При загрузке статьи, должен происходить её анализ, после чего в интерфейсе, выдаваться результат о том:

* Какие ключевые слова можно было бы добавить/убрать;
* К какой тематике относится статья;
* Что можно было бы добавить/исправить.

При этом нужна возможность ручной корректировки:

* Ключевых слов статьи;
* Тематики статьи.

При исправлении пользователем, предложенного ИС результата – должна происходить корректировка весов, и обучение нейронной сети с использованием одного из алгоритмов машинного обучения.

При первичном обучении сети, необходимы примеры статей.

Было принято решение использовать статьи МГОТУ прошлых лет.

Так, же было принято решение, о необходимости примеров статей из сторонних источников, таких как:

* [Elibrary] : <https://www.elibrary.ru/>

Но, необходим доступ, для использования API загрузки статей.

При обучении, возможно применение таких технологий, как языковая модель “Bert”. Представляющий собой метод обработки естественного языка, основанный на использовании нейронной сети.

Была начата разработка первичной основы приложения в формате веб приложения Java/Spring и необходимого интерфейса к нему. Задача на реализацию:

Разработать функционал, позволяющий осуществлять загрузку статей, и выделения из неё нужных частей таких как:

* Ключевые слова;
* Основной текст.

Так же осуществлять сохранение выделенных ключевых слов и текста в базу данных.

Для программной обработки теста загруженных статей, используется Apache POI.