**Доклад «REALEC: метаданные и сайт»**

**Катя Герасименко, Ира Пантелеева**

REALEC метаданные

Метаданные очень ценны для корпуса, поэтому их необходимо хранить максимально полно. Но этика требует анонимизации данных, но просто так забывать имена авторов работ не хочется.

Что хочется иметь:

* сторонние исследователи смотрят / скачивают корпус – имен нет.
* аннотаторы могут вносить и редактировать имена, у пользователей с логином доступ к ним есть.

Это достигается хранением имен и остальных метаданных в разных местах.

Сейчас это устроено следующим образом:

* все метаданные хранятся при текстах в json-файлах.
* имена хранятся в недоступном для скачивания из интерфейса месте в общем для всего корпуса json-файле.

Незалогиненный пользователь видит все метаданные, кроме имени, без возможности редактирования и скачать может только первый json-файл.

Залогиненный пользователь видит форму для редактирования, где есть поле для имени.

С такой системой необходимо следить за заменой путей в json-файле с именами при перемещении файла и систематически делать резервную копию файла с именами.

REALEC сайт

Сайт является важной составляющей корпуса, так как по внешнему виду сайта пользователи могут решить, заходить ли им дальше в корпус или нет. Мы создали удобный интерфейс сайта, чтобы любой мог зайти найти нужную ему информацию. НА главной странице добавлена обновляемая статистика, которая показывает достоинство корпуса REALEC. В докладе будут описаны способы и методы создания такой информации на сайте.

Кроме того, будет сказано, какие изменения произошли в корпусе:

1. добавление новых текстов и папок, отличающихся от предыдущих типом эссе
2. Прчиина выбора парсера UDPipe для частеречной разметки

После доклада хочется посоветоваться, как следует решить проблемы, связанные с сайтом.