Разметка кибербуллинга украинского языка с помощью MAE

1. Описание задачи

Требуется провести разметку лексики, обозначающей кибербуллинг (в широком понимании), по классам.

1. Инструмент MAE
   1. Системные требования[[1]](#footnote-1)

Последняя версия JAVA (не ниже Java 8).

* 1. Процесс открытия документа
     1. Открыть папку с файлом mae.jar. Запустить командную строку (напечатать cmd в адресной строке). В командной строке выполнить команду java -jar mae.jar



* + 1. Должно появиться следующее окно ();

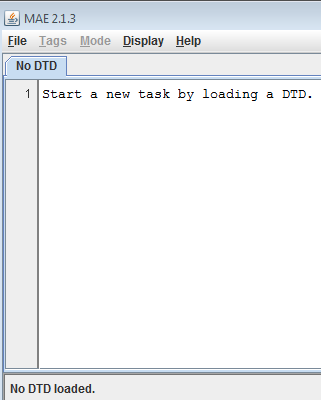


Рисунок 1. Начальная страница MAE

* + 1. Загрузить DTD-файл (File->New task definition ИЛИ Ctrl+N). В открывшемся окне выбрать файл cyberbullying.dtd ()

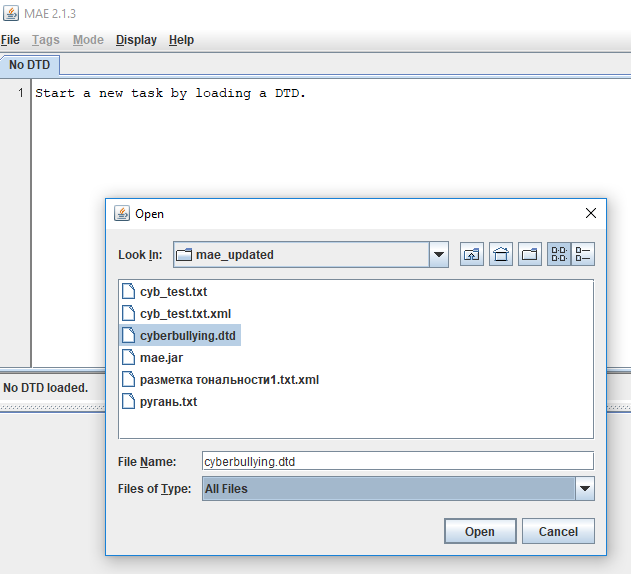


Рисунок 2. Выбор схемы анализа

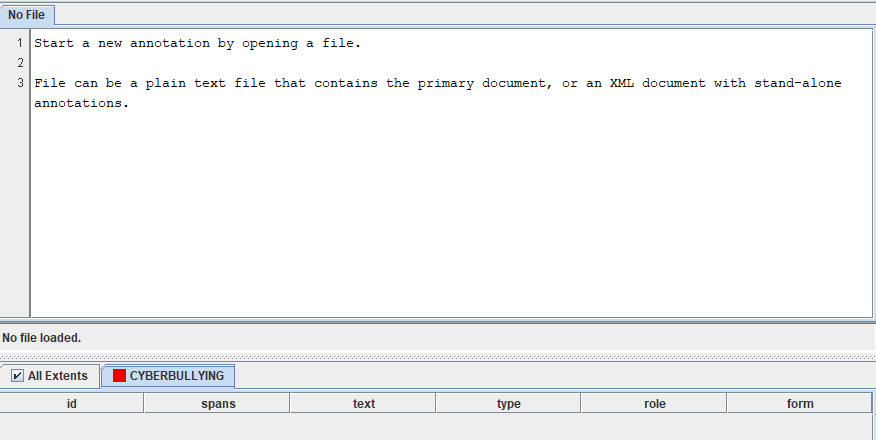
* + 1. Внизу должна появиться схема разметки (). После этого надо выбрать документ для загрузки (File->Open document ИЛИ Ctrl+O). Выбрать документ для разметки, например, файл "cyb\_test.txt”. 

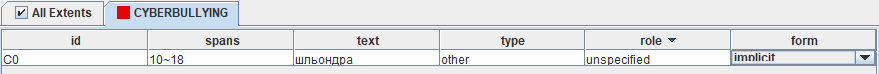
Рисунок 3. Окно успешной загрузки dtd-файла

* 1. Процесс разметки документа

При первом открытии файла разметки следует выбирать файл в формате txt, в дальнейшем — xml с сохраненной разметкой.

* + 1. Разметка проводится для отдельных слов.

Для того, чтобы добавить разметку какому-либо оценочному слову, достаточно дважды щелкнуть на него левой кнопкой мышки. После этого щелчком правой кнопки мышки необходимо выбрать требуемый расширяемый тег "CYBERBULLYING". После выбора данного тега в нижней части страницы появляются редактируемые поля.



* + 1. Разметка проводится по 3 параметрам:
       1. Тип кибербуллинга (type)
       2. Значимость слова в общей оценке (role)
       3. Форма оценки (form)
    2. Расшифровка классов:

Типы кибербуллинга:

* religion — по религиозному признаку
* ideology — по идеологическому признаку (фашизм, нацизм, коммунизм, либерализм и т.п.)
* political\_views — на основании политических взглядов (приверженность конкретной партии, поддержка конкретного кандидата)
* ethnicity — по национальному признаку
* homophobia — по признаку нетрадиционной сексуальной ориентации
* behaviour — по поведению и/или роду деятельности
* appearance — по внешнему виду
* disability — по умственных и/или физическим свойствам
* dehumanization — сравнение с животными или объектами неживой природы
* defecation — сравнение с продуктами жизнедеятельности человека (на всякий случай, привожу определение из Википедии — <https://ru.wikipedia.org/wiki/%D0%94%D0%B5%D1%84%D0%B5%D0%BA%D0%B0%D1%86%D0%B8%D1%8F>)
* other — всё, что не попало в предыдущие категории.

Значимость:

* main — основное слово, по которому определяется тип кибербуллинга
* modifier — слово, вносящее уточняющий и/или дополнительный оттенок
* unspecified — умолчательное значение признака.

Форма оценки:

* explicit — явновыраженная
* implicit (по умолчанию) — подразумеваемая.
  1. Проверка получившегося результата
     1. Результат разметки по каждому отдельному типу можно увидеть в соответствующей вкладке, результат по всем типам можно увидеть в поле «All Extents» ().

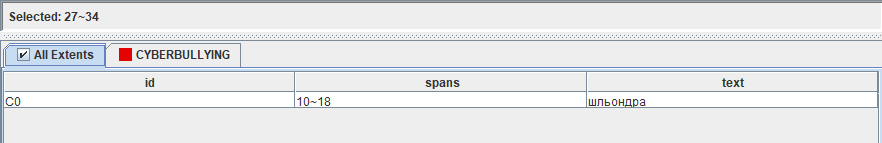


Рисунок 4. Проверка всех типов разметки

* + 1. Вариант 2. Список объектов по классам можно просмотреть в сохраненном xml-файле разметки, разметка добавляется внизу документа.
  1. Сохранение разметки

Сохранить документ можно двумя способами:

* + 1. Комбинацией клавиш Ctrl+S
    2. File->Save annotation as xml

1. FAQ
   1. Размер кегля можно изменять во вкладке «Display».
   2. Чтобы отключить подсветку всех классов размеченных элементов, надо снять «галочку» в поле «All Extents», для каждого поля в отдельности ту же процедуру можно провести, нажимая на цветовое поле.

1. При возникновении проблем с открытием/сохранением файла, просьба переустановить JAVA [↑](#footnote-ref-1)