

Первое информационное письмо

Семинар по оценке алгоритмов поиска заимствований в текстах

Уважаемые коллеги!

В 2016-2017 гг. планируется организовать семинар по оценке алгоритмов поиска заимствований в текстах PlagEvalRus. Организаторы приглашают всех заинтересованных заполнить предварительную заявку на сайте: <https://goo.gl/forms/8BK7ScHODmLcCfOF3>.

Срок подачи заявок: 15 октября, 2016 г.

Официальная страница семинара: <http://ru-eval.ru/plageval>

Контактный адрес: plagevalrus@gmail.com

В ходе семинара планируется решение задачи поиска внешних заимствований (External Plagiarism Detection). Тематика семинара смещена в сторону поиска заимствований в научных текстах (академический плагиат). Предлагаются следующие дорожки (точный набор зависит от количества участников):

1. Дословные заимствования. Copy and paste (c&p) plagiarism detection.
2. Заимствования с парафразами. Paraphrased plagiarism detection.
3. Переводные заимствования с английского на русский. Cross-language plagiarism detection, English-Russian
4. Переводные заимствования с русского на украинский. Cross-language plagiarism detection, Russian-Ukrainian

Каждая дорожка представляет собой поисковое задание: в заданном тексте необходимо найти заимствованные фрагменты и для них указать тексты-источники заимствований из заданной коллекции источников. Участникам предоставляются коллекции текстов источников и обучающие данные по каждой дорожке. Участникам выдаётся также набор тестовых заданий по каждой дорожке. Участники должны предоставить ответы на задания. Полученные ответы сравниваются с эталонными ответами, созданными организаторами. Заимствование считается успешно найденным, если фрагмент текста, найденный участником, вложен во фрагмент эталонного ответа или совпадает с ним. Найденные участниками заимствования, которые не указаны в эталонных ответах организаторов на контрольные задания, не будут учитываться при оценке.

Примерный график проведения

- Сентябрь-октябрь: сбор заявок и обсуждение условий тестирования
- Ноябрь-декабрь: предоставление обучающих данных и подготовка участников к тестированию
- Январь: прогоны тестовых заданий и предоставление результатов организаторам
- Февраль-март: оценка результатов
- Апрель-май: подготовка итогового отчета
- Июнь: подведение итогов (предположительно на конференции Диалог)

Список организаторов:

Иван Смирнов (Институт системного анализа ФИЦ ИУ РАН, Москва), Михаил Копотев (Хельсинкский университет, Финляндия), Андрей Кутузов (Университет Осло, Норвегия), Илья Соченков (Институт системного анализа ФИЦ ИУ РАН, Москва), Ольга Ляшевская (Высшая школа экономики, Москва), Рита Кузнецова (компания Антиплагиат.ру), Олег Бахтеев (компания Антиплагиат.ру), Любовь Иванова (Высшая школа экономики, Москва, секретарь семинара).

Подробные правила и ход проведения тестирования планируется разослать всем заинтересованным участникам для предварительного обсуждения. Приглашаем всех потенциальных участников заполнить анкету-заявку до 15 октября по адресу: <https://goo.gl/forms/8BK7ScHODmLcCfOF3>.

Вопросы и дополнительную информацию можно получить по адресу: plagevalrus@gmail.com
Официальная страница семинара: <http://ru-eval.ru/plageval>

С уважением,
организаторы семинара