Семинар по оценке алгоритмов поиска заимствований в текстах

В 2016-2017 гг. планируется организовать семинар по оценке алгоритмов поиска заимствований в текстах PlagEvalRus. Организаторы готовы включить участников на любом этапе до начала соревнования (декабрь 2016).
Контактный адрес: plagevalrus@gmail.com

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

1. Определение источника заимствования
2. Дословные заимствования: определение заимствованного фрагмента
3. Заимствования с парафразами: определение парафразированного фрагмента

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.

Ход тестирования и оценка результатов
Перед тестированием участникам выдаётся набор заданий по каждой дорожке. Полученные ответы сравниваются с эталонными ответами, созданными организаторами. Заимствование считается успешно найденным, если фрагмент, 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), Олег Бахтеев (компания Антиплагиат), Любовь Иванова (Высшая школа экономики, Москва, секретарь семинара).

Дополнительную информацию можно получить по адресу: plagevalrus@gmail.com.

С уважением,
организаторы семинара