

Первое информационное письмо SpellRuEval

Уважаемые коллеги! В этом году в рамках Dialogue Evaluation впервые планируется провести соревнование по автоматическому исправлению опечаток.

Задача исправления опечаток и нормализации текста является одной из наиболее важных в вычислительной лингвистике. Она значима как сама по себе, так и как промежуточный этап при информационном поиске, автоматической морфологической разметке, извлечении информации из текста, а также в других задачах. Особенно актуальной данная задача является при автоматической обработке текстов из сети Интернет, прежде всего социальных медиа. В последние годы в англоязычной литературе появилось большое количество статей, посвящённых этой проблеме, в которых предлагаются различные подходы, основанные как на словарных, так и на статистических методах, а также на их комбинации. Вместе с тем для русского языка практически отсутствуют как исследования по данному вопросу, выполненные вне поисковых компаний, так и наборы данных, на которых исследователи могли бы тестировать свои подходы. В связи с этим тестирование существующих систем исправления опечаток, а также сравнение различных методологий представляется весьма полезным и интересным. Набор данных, использованный в соревновании, мог бы впоследствии оказаться полезным для разработчиков систем исправления опечаток в качестве эталонной тестовой выборки.

Приглашаем всех желающих принять участие в независимом тестировании систем по исправлению опечаток. В качестве участников мы будем рады видеть как исследователей, уже давно работающих в данной области, так и новичков. В рамках соревнования будет предложена одна дорожка следующего содержания:

Дорожка по исправлению опечаток

Участникам будет предложена выборка предложений, взятых из текстов в социальных сетях, опубликованных в 2010–2015 гг. Требуется исправить все опечатки, содержащиеся в данных предложениях.

Ближайшие действия

В настоящее время нам важно оценить количество потенциальных участников. Если вы рассматриваете возможность принять участие в данном

соревновании, заполните **электронную заявку**. Вы также можете написать письмо Алексею Сорокину по адресу alexey.sorokin@list.ru. Помимо заявок, мы будем рады любым конструктивным комментариям и предложениям по проведению дорожки, а также оценке результатов и другим связанным вопросам.

Предварительные даты

30.11 Первое информационное письмо.

20.12 Публикация обучающей коллекции.

05.02 Публикация тестовой коллекции.

10.02 Приём результатов прогонов на тестовой коллекции.

15.02 Рассылка результатов участникам.

25.02 Подача статьи на конференцию “Диалог”

С уважением, оргкомитет SpellRuEval

Над организацией работает инициативная группа в составе:

Алексей Байтин (Яндекс)

Ирина Галинская (Яндекс)

Евгений Инденбом (АВВУУ)

Елена Рыкунова (МФТИ)

Алексей Сорокин (МГУ, МФТИ)

Татьяна Шаврина (МГУ)