

Российский семинар по оценке методов информационного поиска

Апрель 2010 - Октябрь 2010
<http://romip.ru>

Приглашение к участию

К участию приглашаются все заинтересованные лица - как создатели поисковых систем, так и исследователи, занимающиеся проблемами информационного поиска. В 2010 году очная часть РОМИП будет проводиться совместно с двенадцатой всероссийской научной конференцией RCDL'2010 (Казань, 15-19 октября 2010).

Впервые в программе семинара - дорожка по постоению текстовых меток для изображений и свободная дорожка, в рамках которой Вы можете делать любые исследования на основе материалов РОМИП.

Для студентов и аспирантов предусмотрены льготные условия участия. Участие в РОМИП также является преимуществом при отборе участников летней школы RuSSIR'2010 (<http://romip.ru/russir2010>, Воронеж, 13-18 сентября 2010).

Рассматриваемые задачи

Предварительно, программа семинара в 2010 году будет состоять из следующих секций (дорожек), каждая из которых посвящена оценке методов решения одной из задач информационного поиска:

- классическая задача поиска по запросу
 - по коллекции нормативно-правовых документов
 - по Веб-коллекции
 - по смешанной коллекции
- тематическая классификация
 - нормативно-правовых документов
 - Веб-сайтов
 - Веб-страниц
- фактографический поиск по новостной коллекции
- кластеризация новостного потока
- контекстно-зависимое аннотирование текстовых документов
- поиск похожих документов по документу образцу или фрагменту текста
- вопросно-ответный поиск
- поиск нечетких дубликатов в коллекции изображений
- поиск изображений по визуальному подобию
- построение текстовых меток для изображений
- свободная дорожка (произвольное исследование на основе материалов РОМИП)

Окончательный список будет сформирован по результатам обсуждения в [форуме](#) и предпочтений заявившихся участников. Подробные описания дорожек и правила проведения обсуждаются в [списке рассылки](#) РОМИП и будут опубликованы на сайте семинара.

Наборы данных

При проведении оценки используются следующие коллекции:

- Веб коллекция [Narod.Ru](#) (700.000 документов)
- Веб коллекция [BY.web](#) (1.500.000 документов)
- Коллекция нормативных документов (300.000 документов)
- [Новостная коллекция](#) (32.000 сообщений, 25 источников)
- Веб коллекция [KM.RU](#) (3.000.000 документов)
- Веб коллекция [DMOZ](#) (300.000 документов)
- [Коллекция фотографий](#) (20.000+ изображений)
- [Коллекция нечетких дубликатов изображений](#) (37.000+ изображений)

Процедура и условия участия

Все участники подписывают лицензионные соглашения об использовании коллекций РОМИП исключительно для исследовательских (некоммерческих) целей в рамках проекта РОМИП.

Участник, получив набор данных и заданий, в оговоренные сроки предоставляет результат(ы) выполнения заданий своей системы для их независимой оценки, которая проводится с соблюдением анонимности источника результата. Поощряется также и решение модифицированных задач на основе используемых наборов данных.

Участник обязуется предоставить статью с описанием общих принципов использованного подхода и полученных результатов для публикации в трудах семинара и сделать доклад на очной части семинара.

Участник также обязуется компенсировать часть затрат на распространение тестовых корпусов и оценку результатов, пропорционально степени его участия. Компенсации могут производиться как путем оплаты оргвзносов, так и путем выделения трудовых ресурсов для проведения оценки. Студенты и аспиранты могут получить грант на 50% компенсацию затрат на оценку первых двух предоставленных прогонов.

Результаты тестирования предназначены для использования исключительно в исследовательских целях и не могут быть использованы в маркетинговых или коммерческих целях.

Очный семинар

В 2010 году предполагается совмещение очной части РОМИП со всероссийской конференцией по электронным библиотекам ([RCDL'2010](#)). По предварительному плану очная часть РОМИП пройдет в Казани, в один из дней с 15 по 19 октября 2010 года.

Правила подачи заявки

Заявки принимаются в [электронном виде](#) на сайте семинара (<http://romip.ru>).

Важные даты

Приём заявок на участие: **до 10 мая**

Распространение заданий: до 20 мая

Предоставление результатов выполнения заданий: с 20 июня по 10 июля (по графику)

Результаты оценки: 1 - 30 августа

Тексты докладов: 20 сентября

Очная встреча: 15-19 октября

О РОМИП

РОМИП - это инициатива по проведению независимой оценки методов информационного поиска, ориентированных на работу с русскоязычной информацией.

Методология организации семинара во многом схожа с известными мировыми аналогами - американским TREC, европейским CLEF и японским NTCIR.

Отличительной чертой РОМИП является использование русскоязычных коллекций и активное участие исследователей из России.

Материалы семинаров РОМИП 2003-2009 годов свободно доступны на сайте семинара.

Цели и аудитория

Основными целями Российского семинара по Оценке Методов Информационного Поиска являются:

- создание и развитие информационных ресурсов, обеспечивающих исследования в области информационного поиска;
- проведение независимой оценки методов информационного поиска, ориентированных на работу с русскоязычной информацией;
- формирование среды для исследования проблем информационного поиска на актуальных для российского пользователя задачах;
- формирование требований к оформлению текстовых коллекций для тестирования;
- формирование "правил игры" - этических норм представления и использования результатов.

Дополнительными целями проекта являются:

- консолидация сообщества российских исследователей и разработчиков, занимающихся информационным поиском
- сотрудничество с зарубежными инициативами по тестированию систем информационного поиска
- расширение контактов с зарубежными исследователями и разработчиками, занимающимися многоязычным информационным поиском
- привлечение широкого внимания к проблемам и перспективам информационного поиска и интенсификация исследований в этой области в России
- создание предпосылок и стимулов для дальнейшего развития участвующих систем с целью совершенствования качества поиска.

Оргкомитет

- Михаил Агеев (МГУ, Москва)
- Александр Антонов (Галактика, Москва)
- Павел Браславский (Яндекс, УГТУ-УПИ, Екатеринбург)
- Наталья Васильева (HP Labs, Санкт-Петербург)
- Максим Губин (Facebook, USA)

- Борис Добров (УИС РОССИЯ, Москва)
- Михаил Костин (Mail.Ru, Москва)
- Игорь Кураленок (Яндекс, ПМПУ СПбГУ, Санкт-Петербург)
- Игорь Некрестьянов (Oracle Corporation, USA)
- Марина Некрестьянова (NebuAd, Санкт-Петербург)
- Владимир Плешко (RCO, Москва)
- Илья Сегалович (Яндекс, Москва)
- Влад Шабанов (Вертикальный поиск, Москва)
- Сергей Татевосян (КМ онлайн, Москва)

Дополнительная информация

Веб сайт семинара - <http://romip.ru>.

Основным средством общения и координации в РОМИП является электронная почта. Для получения оперативной информации и обсуждения вопросов, связанных с проведением семинара, приглашаем подписаться на рассылку РОМИП: <http://www.yahoogroups.com/group/romip>.

Контактный адрес оргкомитета: romip@romip.ru