

РОМИП'2009: отчет организаторов

© И. Некрестьянов, М. Некрестьянова

romip@romip.ru

Аннотация

В статье описаны детали организации РОМИП'2009 – дорожки, коллекции, процедуры оценки результатов и другие аспекты проведения семинара. Основное внимание уделено особенностям РОМИП'2009. Подробности о принципах РОМИП и базовых подходах к оценке можно найти в трудах РОМИП прошлых лет [1], где они неоднократно подробно описывались.

1. Введение

В 2009 в программе семинара не появилось новых дорожек и новых коллекций, да и в целом это был непростой год для РОМИП. Тем не менее семинар в 2009 году продолжил развиваться:

- Увеличилось число заявок и участников. В этом году до финиша дошло 16 участников из 20 заявок (и это несмотря на отмену новостной дорожки организаторами).
- 7 заявок на участие поступило от новых участников, и 5 из них дошли до финиша.
- Увеличился объем оценки и число ассессоров, принимавших участие в работе. Всего оценка заняла около 3000 человекочасов, и над ней работало 23 человека.

Подробная информация о заявках и полученных результатах приведена в таблице 1. Сводная статистика о наборе коллекций РОМИП приведена в таблице 2.

Не обошлось и без накладок:

- По решению организаторов были отменены дорожки по кластеризации новостного потока и фактографического поиска по новостной коллекции, несмотря на наличие желающих принять в них участие. Нам не хватило ресурсов на подготовку

инструмента для оценки и формирование правил проведения этих дорожек.

- Дорожки поиска похожих документов и поиск по смешанной коллекции были отменены в связи со сходом заявившихся участников.
- Отменена дорожка вопросно-ответного поиска, поскольку результаты сдал единственный участник.
- Задержка графика предоставления результатов оценки участникам. В частности, оценка по дорожке контекстно-зависимого аннотирования не завершена к моменту публикации трудов.
- Технические накладки при проведении оценки нескольких дорожек (поиск по нормативной коллекции, поиск по KM.RU), что внесло дополнительный шум в результаты.

Во многом это объясняется ограниченными временными ресурсами организаторов и сдвигом графика РОМИП в 2009, но в целом процесс есть куда улучшать и организационно, и технологически. В частности, инструменты оценки, используемые для ряда “старых” дорожек семинара, были созданы несколько лет назад (в очень сжатые сроки). Их стоило бы переделать с использованием современных технологий и обновленного понимания того, как должен быть организован процесс оценки (добровольцы, ау!).

Из других особенностей семинара в 2009 году отметим, что в 2009 году:

- в программе семинара успешно закрепились новые дорожки, посвященные задачам поиска изображений;
- значительно больший объем оценки для дорожки по поиску *Ву.Веб*;
- более активная реакция участников РОМИП на процедуры и результаты оценки.
- ряд участников в этом году попробовали поучаствовать в оценке самостоятельно. В оргкомитет поступила масса критики, советов, предложений и сообщений об ошибках.

В последующих разделах мы подробно расскажем об особенностях проведения дорожек РОМИП в этом году.

Дорожка	Заявившихся участников	Предоставивших результаты	Общее число прогонов
Поиск по Ву.Web	10	8	17
Поиск по КМ.RU	9	7	12
Поиск по нормативной коллекции	8	7	18
Классификация нормативных док.	4	2	2
Классификация Веб сайтов	3	2	4
Классификация Веб страниц	4	3	6
Контекстно-зависимое аннотирование	3	3	7
Поиск по визуальному подобию	2	2	4
Поиск нечетких дубликатов изображений	4	4	4

Таблица 1. Сводная статистика о РОМИП'2009

2. Текстовый поиск

В 2009 году в программе РОМИП было запланировано 5 дорожек для задач поиска по текстовой коллекции. Эти дорожки присутствовали в программе РОМИП и в прошлые годы:

- Классическая задача поиска по запросу по:
 - коллекции нормативно-правовых документов;
 - Веб коллекции ВУ.Web;

- Веб коллекции КМ;
- смешанной коллекции.
- Поиск похожих документов по документу-образцу или фрагменту текста.

Коллекция	Состав	Размер	Предоставлена
Narod.Ru 2003	Веб-сайты из домена narod.ru	728 000 док. 22 000 сайтов	Яндекс
Legal 2004	Законодательство РФ	60 000 док.	Кодекс
DMOZ 2003	Веб-сайты из русской части DMOZ	300 000 док. 2087 сайтов	Рамблер
News 2006	Все новости за три периода из 17 источников	31 500 док. 75 Мб	Яндекс
By.Web 2007	страницы домена .by из индекса Яндекс (май 2007)	1 524 676 док. 8 Гб	Яндекс
КМ.RU 2007	~90% от объема www.km.ru на май 2007 (57 сайтов)	3 010 455 док. 13.7 Гб	КМ Онлайн
Legal 2007	Законодательство РФ, Москвы и Санкт-Петербурга (декабрь 2006)	300 000 док. 1.7 Гб	Кодекс
Flickr 2008	подмножество www.flickr.com	20 000 фот.	Flickr
ImageDupl 2008	стоп-кадры из 15 часов видеоматериала	37 800 изобр.	Оргкомитет РОМИП

Таблица 2. Набор коллекций РОМИП

Фактически были проведены только три первые дорожки. Дорожки по поиску похожих документов и поиску по смешанной коллекции были отменены в связи со сходом заявившихся участников.

2.1 Поиск по коллекции нормативных документов

Правила дорожки не изменились по сравнению с 2008 годом. Поиск производился по коллекции нормативных документов Законодательства Российской Федерации, Москвы и Санкт-Петербурга по состоянию на 2 неделю декабря 2006 года. Общее число документов в коллекции – 300 тысяч. Многие документы представляют собой редакции одних и тех же документов, что с точки зрения классического IR выглядит как большое количество нечетких дубликатов.

Задание – 14974 запросов, отобранных из лога поисковой системы Кодекс в декабре 2006 года. Из них оценивалось - 75. Общее число размеченных запросов для этой коллекции - 220.

Как и в большинстве прошлых циклов РОМИП (за исключением 2004 и 2006), оценка результатов дорожки поиска по нормативно-правовой коллекции производилась экспертами с юридическим образованием. Но, в отличие от предыдущих лет, в 2009 году оценка не аутсорсилась, а оргкомитет работал с ассессорами напрямую.

В оценке участвовало 3 ассессора с высшим юридическим образованием (один ассессор – кандидат юридических наук) и многолетней практикой работы по специальности. Все ассессоры - опытные пользователи ПК и юридических систем.

Как и в 2008 году, глубина пула для этой дорожки составила 35 документов, для каждой пары документ-запрос была собрана одна оценка (на большее не хватило времени экспертов). Общее число оцененных пар документ-запрос - 11824.

Отбор запросов для оценки производился следующим образом. Каждому эксперту был предоставлен свой список в 2000 случайно отобранных ранее не оцениваемых запросов (без какой-либо фильтрации). Списки ассессоров не пересекались. Необходимо было отобрать 25 запросов, руководствуясь следующими принципами:

- не выбирать запросы с опечатками;
- выбирать запросы, по которым эксперту понятно, что искали;
- 10 запросов на поиск с каким-то упоминанием номера документа (статья такая-то и т.д.), предпочтительно разных,

чтобы нельзя было выделить один и тот же шаблон, и менялся бы тип информационной потребности. Например:

- приказ ФТС России от 29 ноября 2006 года N 1252
 - Комментарий ст.14.5 КоАП РФ
 - Форма КС-2
- 15 запросов, где смысл можно понять из текста (не требуется применение шаблонов и т.п., чтобы понять, определить правильный ответ)

Для всех отобранных запросов эксперты должны были составить расширенные описания, которыми они в дальнейшем руководствовались при оценке.

По техническим причинам (неправильно сформированные идентификаторы документов) ответы одного из участников не были учтены при построении пулов. После обнаружения этой проблемы недостающие документы были дооценены теми же ассессорами. Однако таким образом оценка одного задания была разнесена во времени, что могло сказаться на качестве оценки.

В отличие от предыдущих лет в этом году ассессоры руководствовались более формальной инструкцией по оценке (см. Приложение D). При этом, как и в предыдущие циклы, мы предполагали, что оценка экспертов в данной прикладной области более аккуратна, чем, например, оценка в дорожках поиска для Веб. Этим, в частности, было обусловлено решение об использовании одной оценки.

В процессе анализа результатов участниками было обнаружено расхождение в понимании задачи ассессорами и участниками дорожки, которое не было регламентировано правилами. В частности, ассессоры предполагали, что они ищут ответ для “реальной” юридической проблемы, а участники решали задачу текстового поиска. В результате, например, зачастую участники считали все редакции одного документа релевантными, а ассессоры - нет.

Мы провели предварительный анализ этой проблемы, используя группы “очень близких по тексту” документов среди оценивавшихся, которые построил и предоставил оргкомитету Михаил Агеев. Всего 108 кластеров, содержащих 535 документов (кластеры обнаружены для 39 запросов из 75 оценивавшихся).

Эксперт (с ученой степенью и стажем работы), который не участвовал в оценке этой дорожки, выставил свои оценки для этих кластеров (руководствуясь только текстом запроса и содержимым документов, без расширенного описания). Относительно часто (хотя

мы пока не располагаем точной численной оценкой) его мнение расходилось со мнением других ассессоров, что подтверждает субъективность даже экспертных оценок.

Среди причин, по которым только некоторые из схожих документов были признаны релевантными, эксперт отметил, например:

- Документы недействительны (на конец 2006 года), и существует их обновленная редакция, или в документах явно сказано, что они утратили силу/отменены.
- Есть “другой” новый закон/нормативный акт, который регламентирует этот вопрос. Этот факт отсутствует в результатах, но ассессор “знает” о его существовании.
- “Проект” закона бесполезен, нужен окончательный принятый текст.

Мы планируем закончить верификацию кластеров и использовать этот опыт для:

- Уточнения постановки задачи для участников и ассессоров, чтобы снизить расхождение в их понимании;
- Оценить влияние “расхождения” на выводы о превосходстве методов (устойчивость результата).

2.2 Поиск по Веб коллекциям

Как и в 2008 году, в программе было две дорожки поиска по Веб-коллекции с общим набором заданий:

- Поиск по коллекции **Vu.Web**
- Поиск по коллекции **KM.RU**

Набор заданий не изменился с прошлого года и состоял из 29231 запросов, полученных из журналов поисковых систем. При оценке результатов для каждой дорожки использовались только запросы из журнала запросов, соответствующего данной коллекции.

Ассессоры использовали один и тот же инструмент оценки и руководствовались единой инструкцией, регламентирующей описание информационных потребностей и примеры оценок (см. Приложение В). Единственным отличием инструкции 2009 года от той, что использовалась в 2008, является явное указание, что задания поиска по **vu.web** надо оценивать исходя из позиции пользователя, находящегося на территории Белоруссии:

*“При оценке заданий для коллекции **VU.WEB**, пожалуйста, представьте себя на месте человека, проживающего в*

Белоруссии. То есть учитывайте географическую специфику запросов и ответов, если она есть. Например, для запроса "купить носки" - релевантна информация о том, как это сделать в магазинах Белоруссии или в Интернете с доставкой. Продажа носков в Москве - слаборелевантна."

Для каждого оцениваемого задания оценки собирались от двух ассессоров. При этом ассессор мог отказаться от оценки запроса, если задание было ему совершенно непонятным или предлагаемый запрос был порнографическим. Замена ассессора в таком случае не производилась (этой возможностью воспользовались 15 раз).

Перед началом работы над заданием и после его окончания ассессоры заполняли небольшую анкету, характеризующую их понимание задания. Начальная анкета содержит следующие вопросы:

- Вы можете отказаться от оценки этого задания по одной из следующих причин
 - запрос связан с порнографией
 - задание совершенно непонятно
 - не буду отказываться!
- Пожалуйста, отметьте, что из нижеследующего верно для этого задания
 - в тексте запроса есть опечатки
 - текст запроса содержит важные сокращения (которые, вероятно, будут в расшифрованном виде в тексте ответа))
- Насколько вам понятна цель поиска?
 - Плохо понятно
 - В общем, понятно
 - Ясно
- Как вам кажется, что является целью поиска:
 - Нахождение адреса конкретного сайта
 - Получение конкретного ответа на конкретный вопрос
 - Получение ответа на четкий вопрос, для которого нет единого ответа на вопрос
 - Собрать информацию на заданную тему
 - Найти информацию о том, где можно получить/купить какой-то продукт/сервис
 - Поиск ресурса, а не информации, который доступен на странице или по ссылке со страницы
 - Что-то другое

По завершении задания ассессор должен был ответить на следующие вопросы:

- Часто ли у вас возникали сомнения при принятии решения по сравнению с другими заданиями?
 - Чаще, чем обычно
 - Как обычно
 - Реже, чем обычно
- Насколько вы уверены в принятых решениях по сравнению с другими заданиями?
 - Более, чем в среднем
 - Как обычно
 - Меньше, чем в среднем
- Разница между какими классами релевантности вам было сложно определить?
 - "Соответствует" и "скорее соответствует"
 - "Скорее соответствует" и "возможно соответствует"
 - "Возможно соответствует" и "не соответствует"

Особенности и детали оценки для каждой из дорожек описаны в следующих разделах.

2.2.1 Поиск по VY.Web

Как и в 2008 году, для этой дорожки оценивается значительно больше запросов, чем для других поисковых дорожек, но с меньшей глубиной пула. В этом году оценивалось 550 запросов - 500 новых запросов и 50 оценивавшихся в прошлом году. Глубина пула – 20.

Даже при такой небольшой глубине пула трудоемкость оценки этой дорожки очень велика – в этом году собрано 150.000 оценок, что в 6 раз больше, чем для дорожки поиска по KM.RU в 2009, и на 70% больше, чем для поиска для Vу.Web в 2008. Всего на данный момент для коллекции Vу.Web размечено 1060 запросов.

Еще одна особенность этой коллекции – случайный отбор запросов с минимальной фильтрацией (нечитаемая кодировка, порнографические запросы). Всего было отобрано 600 запросов, из которых было отфильтровано около 40. Оценивались первые 500 из оставшихся.

При оценке задания каждый из ассессоров должен был составить свое описание информационной потребности, соответствующей запросу. Первая версия описания составлялась до начала оценки, то есть ассессор видел только текст запроса, но не видел ни одного из оцениваемых документов. Ассессор мог исправить расширенное

описание по окончании выполнения оценки этого запроса, и этой возможностью довольно активно пользовались. Так, в 287 случаях из 1100 ассессоры не создавали никакого расширенного описания до начала оценки, а заполняли его только после завершения оценки. Еще в 214 случаях ассессор уточнял первоначальное описание после завершения оценки.

2.2.2 Поиск по KM.RU

Правила проведения этой дорожки близки к “традиционной/классической” дорожке поиска в РОМИП.

Оценивается относительно небольшое число (50) запросов с относительно большой глубиной пула (50), что позволяет лучше аппроксимировать полноту.

Из случайной выборки запросов вручную отсеивались мусор, опечатки, явно навигационные запросы для ресурсов, которых нет в коллекции (например, одноклассники.ру). Благодаря более строгой фильтрации запросов все запросы были оценены. Всего на данный момент для коллекции KM.RU размечено 170 запросов.

Важным отличием этой дорожки от Vu.Web является составление расширенных описаний первым ассессором и их повторное использование вторым.

Несмотря на “общее” расширенное описание, согласованность оценок для этой дорожки формально ниже, чем для Vu.Web (0.74 против 0.85). Однако это оказалось следствием технической проблемы у части ассессоров, в результате которой они видели частичное содержание для ряда документов. Эта проблема касается одной оценки для 16 запросов, и согласованность оценки на этих запросах – 0.4. Для остальных 34 запросов уровень согласованности 0.9, что выше, чем у Vu.Web, и согласуется с нашими ожиданиями.

Мы планируем переоценить неправильно оцененные задания.

3. Текстовая классификация

Набор дорожек для задач текстовой классификации не изменился с прошлого года, и не было значительных изменений в правилах и инструменте оценки. Основные факты для каждой из дорожек представлены в последующих разделах.

3.1 Нормативных документов

Правила этой дорожки не изменились с прошлого года. Таксономия основана на рубрикаторе, предоставленном компанией Кодекс в 2006 году и состоящем из 720 рубрик. Всего в таксономии упоминаются 29573 документа, но они могут быть отнесены к нескольким рубрикам. Общее число обучающих примеров – 94599.

Оценка была произведена по 75 случайно выбранным и ранее не оценивавшимся категориям методом сравнения с рубрикацией, предоставленной компанией Кодекс вместе с коллекцией (эталонный рубрикатор был составлен вручную). При вычислении оценок НЕ учитывались документы, которые были использованы для обучения по данной категории (и такие документы НЕ включены в итоговую таблицу релевантности).

3.2 Веб сайтов

Постановка задачи и правила проведения дорожки не изменялись с 2007 года (а большинство правил и обучающее множество используются, начиная с 2004). Обучающее множество было создано на основе каталога DMOZ и содержит 247 рубрик и 2116 обучающих примеров (Веб сайтов).

Один и тот же сайт может относиться сразу к нескольким категориям. Ответ системы - упорядоченный список (до 5 категорий) для каждого из классифицируемых сайтов. Сайт может не относиться ни к одной из категорий, и в этом случае идеальным ответом является пустой список назначенных категорий.

Оценка производилась по 20 категориям, по 2-4 категории для 6 категорий первого уровня. Каждое задание оценивали два ассессора.

При проведении оценки крупные сайты (более 200 страниц) передавались ассессорам в сокращенном виде. В сокращенное множество включались только документы, которые в сжатом виде занимали не более 150000 байт. Всего отбиралось 200 документов, так чтобы было хотя бы по одному документу для каждой папки верхнего уровня или скрипта (то есть для каждой строки, получаемой из url путем обрезания по первому символу '/' или '?' после имени сайта).

3.3 Веб страниц

Начиная с 2008 года, в правилах этой дорожки есть важное отличие: ответ должен был состоять из упорядоченных по близости к теме рубрики списков документов для каждой рубрики.

В силу невозможности оценить все отнесенные к категории документы (полные котлы для категорий содержат десятки тысяч документов, и для их оценки необходимо произвести огромный объем работы), оценивается некоторая выборка из результатов каждой системы. Упорядоченность результатов и использование верхушки списка ответов для оценки позволяет значительно повысить долю релевантных документов в котле, что дает лучшее представление о качестве работы систем.

Оценка производилась по 20 (15 в 2008) категориям. Оценивалось только 50 первых (в 2008 году 75) документов для каждой категории из каждого прогона. Поэтому не для всех возвращенных документов из каждого прогона систем есть оценки релевантности, что, безусловно, сказывается на общих оценках систем.

Для получения более полной картины рассчитывались два набора оценок:

- Учитывались все документы, и те документы, для которых нет оценки, считались нерелевантными.
- Учитывались только оцениваемые документы. Документы, для которых нет оценок, не учитывались (как будто система не относил их к оцениваемым категориям).

4. Контекстно-зависимое аннотирование

Набор заданий для этой дорожки был построен на основе таблиц релевантности дорожек по поиску РОМИП 2007 и 2008. Всего было 88497 заданий, соответствующих 757 разным запросам к коллекции нормативных документов и Веб коллекции.

Как и в прошлом году, ответом системы для задания является фрагмент текста не более 300 символов без HTML разметки (в том числе,
 и <p>). В дополнение к аннотации ассессор видит заголовок документа (содержимое тега title, до 100 символов). Размер заголовка не учитывается в размере аннотации.

Опираясь на успешный опыт по организации этой дорожки в 2008 году, мы стремились обеспечить ассессору одновременный доступ ко всем вариантам аннотаций для заданной пары запрос-документ. При оценке также отдавалось предпочтение расширению спектра запросов за счет уменьшения числа документов, чьи аннотации для этого запроса оцениваются. Оценка аннотаций производилось только для документов из Веб-коллекций.

Как и в 2008 году:

- Задания группировались в наборы по запросам. Набор состоит из последовательности заданий, каждое из которых соответствует одному документу.
- Ассессор видел ВСЕ доступные аннотации в случайном порядке (заголовок документа + текст аннотации, обрезанный до 300 символов).
- Для каждой аннотации необходимо было выставить 2 оценки:
 - информативность
 - читабельность
- Ассессор также должен был ответить на два вопроса, характеризующие его понимание о релевантности документа и о роли заголовка:
 - исходя из полученной информации, считаете ли вы, что документ содержит релевантную информацию?
 - приняли ли бы вы такое же решение, используя ТОЛЬКО заголовок документа?

Информативность характеризует, насколько эта аннотация понятна для принятия решения о полезности документа в контексте этого запроса.

Критерий читабельности определял ответ на следующий вопрос: “Аннотации зачастую состоят из обрывков приложений и отдельных словосочетаний. Мешает ли вам это понимать их смысл?”.

При этом мы полагаем, что высокой информативностью и читабельностью могут обладать и аннотации для нерелевантных документов, так как цель аннотации - помочь пользователю принять правильное решение о полезности документа.

Однако в 2009 году у нас было 7 прогонов (а не 3), что потребовало изменений в инструменте оценки. Еще одна мотивация для изменений по опыту прошлого года – путаница, возникающая при оценке в связи с присутствием двух шкал для каждого документа.

На момент публикации этой статьи оценка дорожки не завершена, поэтому мы не можем обоснованно оценить, насколько успешными оказались сделанные изменения. Результаты будут опубликованы на сайте семинара по мере завершения оценки.

5. Поиск изображений по образцу

Повторение дорожки, добавленной в программу РОМИП в 2008 году, – та же коллекция (Flickr, 20000 фотографий), набор заданий и правил, не изменившийся инструмент оценки (см. Рисунок 3).

Дорожка посвящена оценке методов решения задачи поиска по содержанию изображений (content-based image retrieval) на коллекции разнородных фотографий, типичных для персональных непрофессиональных фотоархивов.

Фотографии в коллекции не связаны с какой-либо дополнительной информацией (такой как аннотации, теги или другой контекст).

Участникам необходимо было отобрать изображения, похожие на изображение-образец визуально и семантически с точки зрения человека. Релевантными изображениями считались как глобально похожие, так и обладающие локальным сходством.

Изображения глобально похожи, если на них представлены одинаковые сцены (например, два снимка ночного города). Изображения обладают локальным сходством, если на них представлены похожие объекты на разном фоне.

Глобальное сходство играет решающую роль в случае, когда на изображениях сложно выделить центральный объект (в основном, пейзажные фотографии), в то время как локальное сходство важно для фотографий с явно выраженным объектом съемки (портретное фото, съемка животных).

Для локального сходства не требуется идентичности объектов на различных снимках, объекты должны быть одного вида, одной природы. Так, два портретных изображения разных людей могут быть признаны похожими при наличии некоторого визуального сходства (одна поза, одинаковая длина и цвет волос и т.д.).

Примеры изображений можно найти в инструкции ассессора для этой дорожки в приложении Е.

Как и в прошлом году, оценка производилась во многом аналогично оценке для задач текстового поиска. Для оценки было случайным образом отобрано 250 заданий (то есть изображений образцов). В котлы попало по 20 первых результатов в каждом прогоне. Каждое задание оценивало 2 ассессора, и всего собрано 36000 оценок. Важным отличием этого года являлось наличие формальной инструкции для ассессоров (см. Приложение Е).

Отметим, что при примерно том же объеме оценки в этом году больше релевантных ответов. Так, в прошлом году было 701 и 2282

слаборелевантных документа (AND и OR соответственно), а в 2009 – 1473 и 4403 соответственно.

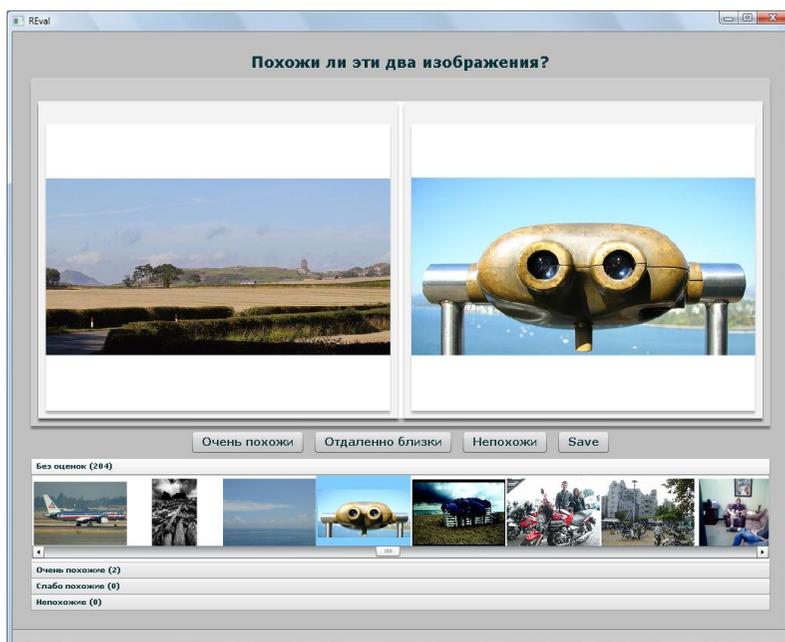


Рисунок 1. Инструмент оценки для дорожки поиска по визуальному подобию.

6. Выявление нечетких дубликатов в коллекции изображений

Эта дорожка посвящена оценке методов поиска нечетких естественных дубликатов в коллекции фотографий и в 2009 году являлась повторением дорожки 2008 года. Как и в прошлом году, оценка производилась одним ассессором, хорошо знакомым с содержимым коллекции.

Дубликатами считаются фотографии одной и той же сцены или объекта, сделанные в разных условиях или разного качества. В частности, дубликатами являются фотографии, снятые в разном масштабе или с разных точек, с различиями в фокусном расстоянии,

освещении, с незначительными изменениями фона (движение волны в море или листьев на дереве).

Система-участник должна определить имеющиеся группы дублей в коллекции. Допускается, что одно изображение входит в несколько различных групп дублей одновременно. Ограничений на размеры групп нет.

Примеры "естественных" дубликатов:



Примеры визуально и/или семантически похожих изображений, не являющихся при этом дублями:





Коллекция состоит из 37800 изображений, полученных путем выборки случайных кадров из 15 часов видеоматериала в нескольких различных разрешениях. Качество изображений варьируется в широких пределах.

Оценка производилась путем выделения кластеров дублей из окрестностей 25 случайных изображений. Из каждой окрестности можно было выделить до 20 кластеров.

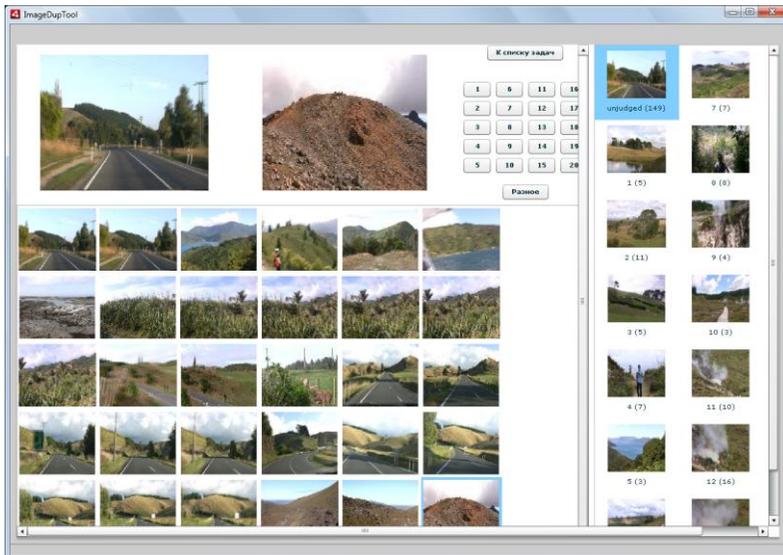


Рисунок 2. Инструмент оценки для дорожки поиска нечетких дубликатов изображений.

Формально окрестность строилась как результат объединения:

- Всех изображений из прогонов участников, которые попали в кластеры с этим изображением.
- Изображений из временной окрестности этого изображения (используя информацию о связи этого изображения с кадрами в исходном видеопотоке).

В оценке принимал участие один ассессор, и всего было получено 466 групп.

Такой подход позволяет вычислить оценки и для прогонов прошлого года, не увеличивая объем работ по оценке. Интересно, что результаты прошлого года выше, чем результаты в этом году (на наборе кластеров выявленных в 2009 году).

Заключение

Несмотря на все трудности, РОМИП развивается – появляются новые участники и увеличивается объем и качество собранных материалов, растет активность участников.

Но есть и проблемы, организационные и технологические, связанные как с ростом семинара, так и медленной ротацией состава активных участников оргкомитета. Надеемся, что в будущем удастся решить хотя бы часть из них.

Мы хотим выразить благодарность всем, кто активно помогал проводить семинар в 2009 году: совершенствовал правила проведения семинара, помогал находить ошибки в инструментах оценки и задавал “неудобные” вопросы, участвовал в оценке результатов, редактировании трудов и обеспечивал поддержку существования НП РОМИП. Без Вашей помощи РОМИП’2009 не состоялся бы. Спасибо!

Литература

- [1] Труды РОМИП онлайн. <http://romip.ru>
- [2] И.Некрестьянов, М. Некрестьянова. РОМИП’2006: отчет организаторов. РОМИП’2006.
- [3] Li Chen, F. W.M. Stentiford. Comparison of Near-Duplicate Image Matching. Visual Media Production, 2006. CVMP 2006, p. 38-42, 2006.

ROMIP 2009: Report from Organizers

Igor Nekrestyanov, Marina Nekrestyanova

This report describes details of ROMIP'2009 evaluation activities from organizers perspective. We briefly describe track rules and focus on specifics of evaluation procedures used during this cycle.