Mail.Ru на РОМИП-2005

© Федоровский Андрей, Костин Михаил,

Проскурин Андрей

Mail.Ru
fedorovsky@corp.mail.ru, kostin@corp.mail.ru,
proskurin@corp.mail.ru

Аннотация

Статья посвящена описанию итогов участия компании Mail.Ru в дорожках информационного поиска и классификации на семинаре РОМИП'2005. Приведено краткое описание архитектуры используемых систем. Особое внимание уделено методам формирования функции релевантности поисковой системы и их влиянию на полученные результаты. Проведены также исследования устойчивости алгоритмов на разных подмножествах запросов.

1. Введение

Целью участия компании Mail.Ru на семинаре РОМИП'2005 являлась в первую очередь обкатка разработанных в компании технологий поиска и классификации информации. В данной работе дано описание использующихся для этой цели алгоритмов и проведена проверка их состоятельности путем сравнения результатов выполнения заданий дорожек РОМИП с показателями, достигнутыми другими участниками.

2. Дорожки информационного поиска (ad hoc)

2.1 Общее описание системы

Поисковая система, используемая для проведения экспериментов по дорожкам, была разработана в компании в рамках

проекта Поиск@Mail.Ru (http://go.mail.ru/) и успешно используется в настоящее время на веб-проектах компании и наших партнеров.

При разработке системы одной из главных целей было сделать ее гибкой и легко адаптируемой при помощи удобного набора настроек к самым разным поисковым задачам, значительно отличающимся друг от друга как по характеру индексируемой коллекции, так и по типичным поисковым потребностям пользователей. С другой стороны, необходимо было обеспечить высокое качество поиска по комплексной коллекции, содержащей различные типы документов, типичным примером которой является веб-коллекция по достаточно большому набору сайтов. Кроме того, важным требованием была высокая производительность, позволяющая использовать систему для поиска по большим объемам данных под высокой пользовательской нагрузкой.

Структура поискового индекса близка к классической, многократно описанной в литературе [3, 4]. Основой являются инверсные списки вхождений слов, используемые для поиска релевантных документов и прямые индексы термов для формирования фрагментов, возвращаемых пользователю (сниппетов). Применяется также ряд техник, позволяющих уменьшить требования к памяти и увеличивающих скорость работы как во время индексации, так и при обслуживании пользовательских запросов.

Для обеспечения высокого качества возвращаемых результатов в первую очередь необходим правильный выбор функции релевантности, определяющей меру соответствия документа запросу. Остановимся на этом более подробно.

2.2 Особенности ранжирования в Поиск@Mail.Ru

При расчете релевантности нами учитывается как частота вхождения в документ единичных слов запроса, так и совместная встречаемость слов и их взаимное положение. В отличие от большинства известных систем, мы используем два различных способа учета взаимного положения слов в документе: совместное вхождение пар слов и нахождение в документе релевантных пассажей. Каждый из этих методов имеет свои достоинства: метод пар слов позволяет качественно обрабатывать не только запросы, являющиеся единым словосочетанием, но и запросы с разной связанностью групп слов внутри запроса. В то же время, наиболее близкий к запросу пассаж лучше других методов позволяет оценить наличие формального соответствия между запросом и документом,

то есть наличие в документе хотя бы простого упоминания объекта, заданного в запросе. При совместном использовании эти два метода, на наш взгляд, хорошо дополняют друг друга и позволяют добиться хорошего качества поиска для максимально широкого круга запросов.

Таким образом, вес документа по запросу в нашей системе складывается из трех составляющих:

$$W = k_f W_f + k_p W_p + k_{ps} W_{ps} \quad (1)$$

где:

W_f – вес документа, вычисленный на основе TF*IDF алгоритма;

W_p – вес документа, вычисленный на основе совместных вхождений в документ пар слов, расположенных рядом в запросе;

W_{ps} – вес наиболее близкого к запросу пассажа документа;

k_f, k_p, k_{ps} – коэффициенты.

Следует также отметить, что для получения ненулевого веса в документе не обязательно должны присутствовать все слова запроса. В число ранжируемых попадают также документы, для которых отношение суммарного IDF слов запроса, встречающихся в них, к суммарному IDF всех слов запроса превышает заданный порог. Такие документы дополнительно «штрафуются» за отсутствующие слова, однако, тем не менее, вес некоторых из них в общем случае может даже превышать вес документов, содержащих все слова запроса.

Рассмотрим подробнее каждый из весов в формуле (1).

2.2.1 TF*IDF вес

Формула, используемая нами для подсчета TF*IDF веса по каждому терму запроса, является модификацией стандартной BM25 формулы [2], и выглядит следующим образом

$$TF * IDF_{term} = \frac{f_{term} \times IDF_{term}}{f_{term} + k_1(b + L(1-b))} \quad (2)$$

где:

f_{term} – вес терма в документе, вычисленный на основе количества вхождений, с учетом ряда дополнительных факторов;

IDF_{term} – обратная частотность терма в коллекции, вычисленная по стандартной логарифмической формуле;

L – нормированная длина документа;

k_1, b – коэффициенты.

Общий TF*IDF вес документа получается суммированием полученных весов по всем термам запроса.

Особенностью применения этой формулы в нашем поиске является то, что для документов, размер которых превышает константу k_2 (соответствующую в стандартной BM25 формуле средней длине документа в коллекции, а в нашей системе задаваемой в настройках) вместо нормирования по длине используется метод разбиения документа на перекрывающиеся фрагменты [1]. Применение этого метода позволяет избежать неоправданного занижения веса длинных документов, в которых имеется небольшой фрагмент с высокой релевантностью.

Фрагменты имеют фиксированный, задаваемый в настройках размер, меньший k_2, и берутся с наложением по всему тексту документа.

Вес каждого из фрагментов по каждому терму запроса оценивается по формуле (2) без нормирования по длине, то есть с $L = 1$.

В результате, выбирается фрагмент документа, имеющий наибольший вес и его вес используется в качестве W_f веса документа в (1).

В качестве особенности, не встречающейся в известных нам работах на эту тему, можно отметить, что в каждый фрагмент нами дополнительно включается небольшой отрезок текста в начале документа, существенно меньший, чем длина фрагмента, так как слова, находящиеся в самом начале длинного документа, часто описывают его содержание в целом.

Для документов, имеющих длину меньшую, чем k_3, используется нормирование по длине по формуле

$$L = \frac{L_w + k_4}{k_3 + k_4} \qquad (3)$$

где:

L_w – длина документа в словах;

k_3, k_4 – коэффициенты, задаваемые в настройках.

Еще одной существенной особенностью TF*IDF ранжирования в нашем поиске является использование достаточно большого значения коэффициента k_1 в формуле (1): для прогонов использовалось значение, значительно б**о**льшее обычно принятых. Наш выбор здесь связан с тем, что небольшое значение этого коэффициента призвано дать преимущество документам с достаточно одинаковой встречаемостью в документе различных слов запроса, что актуально в случае, когда TF*IDF является единственным критерием ранжирования и совместная встречаемость слов никак иначе не учитывается. Поскольку мы учитываем совместную встречаемость слов отдельно, то здесь мы выбрали значение коэффициента, позволяющее дать достаточно высокий вес документам с высокой встречаемостью лишь некоторых (и даже, в частности, одного) слов запроса.

2.2.2 Вес по парам слов

При подсчете этого веса вхождение терма в документ учитывается только в том случае, если оно находится в документе на расстоянии, не превышающим заданное от хотя бы одного из стоящих рядом с ним термов запроса (особая обработка предусмотрена для стоп-слов).

Для прогона по веб-коллекции расстояние было нами выбрано как 2 (рядом или через одно) для случая, когда порядок слов в запросе и документе совпадает и 1 (только рядом) для случая, когда не совпадает.

Соответствующие этому условию вхождения слов обрабатываются по описанному выше TF*IDF алгоритму, отличается только набор коэффициентов.

2.2.3 Вес лучшего пассажа

Под пассажем мы понимаем фрагмент документа, размера, не превышающего заданный, в котором встречаются все термы запроса, либо значительная часть термов запроса, суммарный IDF которых превышает заданное ограничение.

При выборе лучшего пассажа документа основными факторами являются его полнота (наличие всех термов запроса), длина, порядок слов (его совпадение с порядком слов в документе), зона документа (заголовок, выделенный текст, обычный текст), в которой встретился пассаж, близость пассажа к началу документа. Учитывается также ряд дополнительных факторов.

Вес пассажа по каждому из факторов оценивается в баллах на основе специальных для каждого из них правил, после чего веса суммируются. Суммарный вес и будет весом пассажа. Из полученных весов пассажей выбирается максимальный для вычисления общего веса по формуле (1).

2.3 Результаты участия системы в дорожках поиска (*-adhoc)

Мы приняли участие во всех трех дорожках поиска:

- поиск по веб-коллекции (web-adhoc) – 1 прогон;
- поиск по нормативным документам (legal-adhoc) – 1 прогон;
- поиск по смешанной коллекции (mixed-adhoc) – 1 прогон.

Гибкость настроек системы оказалось достаточной, чтобы по каждой из дорожек однозначно выбрать параметры, желаемые для построения гармоничной системы, настроенной на данную коллекцию и, как следствие, ограничиться одним прогоном. Конечно, глобальный оптимум мог и не быть достигнут, однако результаты в каждом случае оказались вполне удовлетворительными.

2.3.1 Web-adhoc

Для дорожки поиска по веб-коллекции (web-adhoc) участникам была предложена достаточно обширная коллекция веб-страниц, представляющая из себя часть сайтов домена narod.ru (более 700000 страниц, 6.3ГБ). По условиям дорожки по этой коллекции необходимо было выполнить большое количество (более 24000) запросов, специально отобранных из поисковых логов. Первые 100 документов из поисковой выдачи считались ответом системы на запрос, упорядоченный по мере убывания значимости документов в выдаче. И коллекция, и список запросов предлагались те же, что и на РОМИП-2004. Для итоговой оценки после получения результатов из всего множества запросов были отобраны 75 (25 из прошлогодних запросов и 50 новых). Достаточно подробное описание способов производимых оценок и стандартных параметров приведено, например, в описании прошлогоднего семинара [5].

Ниже приведена таблица результатов участников дорожки для способа оценки web-adhoc-or-pd50-all (хотя бы одна оценка превышает минимальный порог релевантности; рассматриваются ответы систем, суженные до глубины пула – 50 документов; по

запросам 2004 и 2005 годов) и 11-точечный график TREC по тому же способу оценки.

Прогон	1	2	**Mail.Ru**	4	5	6
Recall	0,4023	0,2471	**0,5443**	0,4548	0,4265	0,4627
Precision(5)	0,3707	0,2000	**0,5840**	0,4853	0,5013	0,5013
Average precision	0,2027	0,0933	**0,3178**	0,2488	0,2401	0,2585
Precision(10)	0,3507	0,2133	**0,5147**	0,4467	0,4373	0,4560
R-precision	0,2733	0,1509	**0,3521**	0,3074	0,2925	0,3138
Precision	0,2701	0,1797	**0,3568**	0,3045	0,2999	0,3231

Таблица 1. Сравнительные результаты оценки web-adhoc-or-pd50-all.

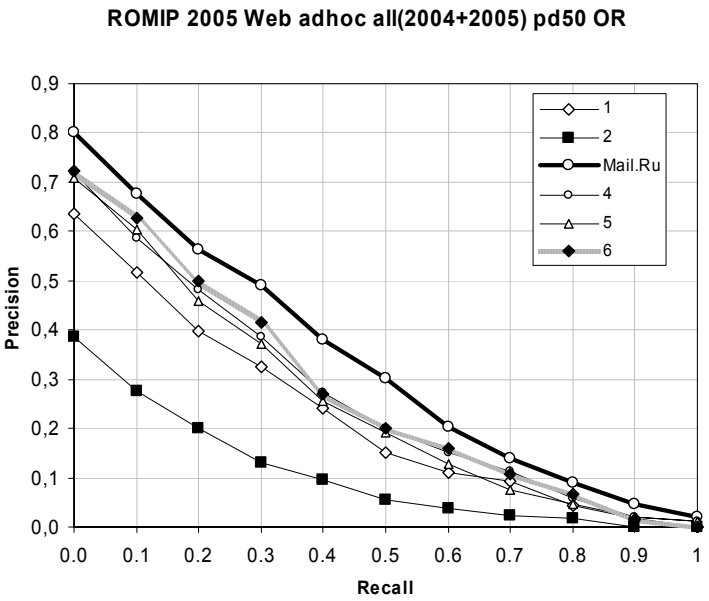


Рисунок 1. 11-точечные графики TREC для участников дорожки веб-поиска с оценкой web-adhoc-or-pd50-all.

Примерно такое же соотношение как параметра Average Precision, так и точек графика TREC между результатами нашей системы и других участников сохраняется и для остальных способов оценки.

В этой таблице – значения метрик системы Поиск@Mail.Ru по различным видам оценок. В первых двух колонках –показатели рассчитываются для всех документов из поисковой выдачи, в последних двух – только для первых 50 (pd50).

	вся выдача		pd50	
	AND	OR	AND	OR
Recall	0,7634	0,6847	0,6309	0,5443
Precision(5)	0,4149	0,5840	0,4149	0,5840
Average precision	0,3330	0,3704	0,3061	0,3178
Precision(10)	0,3179	0,5147	0,3179	0,5147
R-precision	0,3331	0,3826	0,3207	0,3521
Precision	0,1235	0,2460	0,1824	0,3568

Таблица 2. Результаты прогона Поиск@Mail.Ru для оценок вида web-adhoc-*-*-all.

Для исследования устойчивости результатов и поведения нашей системы в сравнении с другими участниками мы разбивали запросы на подгруппы по разным признакам. Для экспериментов мы также выбрали способ оценки web-adhoc-or-pd50-all.

В первую очередь мы решили исследовать зависимость результатов от количества слов в запросе. Однако, поскольку запросы с одинаковым количеством слов могут быть совершенно разного характера, подгруппы запросов не должны быть слишком маленькими, иначе излишне сильны будут случайные возмущения результатов. В связи с небольшим количеством (75) отобранных к оценке запросов и слишком большим квантом изменения в данном методе разбивки удалось выбрать только одно приемлемое множество подгрупп запросов: «2 слова» (29 запросов), «3 слова» (24) и «4 и более слов» (22). Показателем «качества работы» систем мы выбрали Average precision (AvP), как это сделано, например, в поисковых дорожках конференции TREC. Диаграмма зависимости этого параметра от количества слов приведена на рис. 2.

Помимо хорошей устойчивости нашей собственной системы, мы заметили любопытный факт: для участников из прогонов 2 и 5, так же, как и для нашей системы, было характерно **плавное** снижение AvP с ростом количества слов (см. рис. 3).

В то же время у участников 1, 4 и 6 гораздо ярче виден всплеск на двухсловных запросах а на 3-словных, наоборот, происходит резкий спад (см. рис.4). Возможно, это характеризует используемые алгоритмы систем.

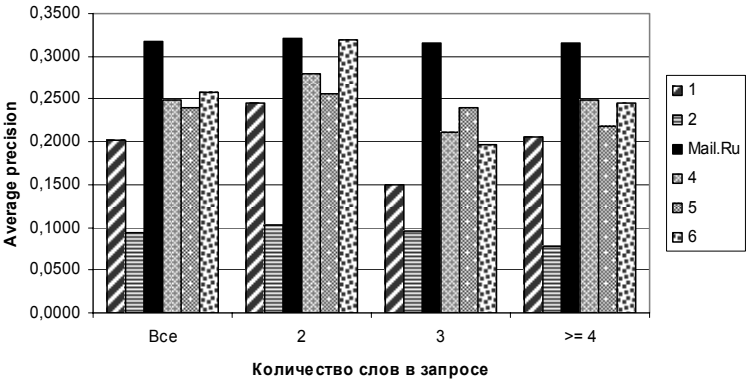


Рисунок 2. Значения AvP для разного количества слов в запросе.

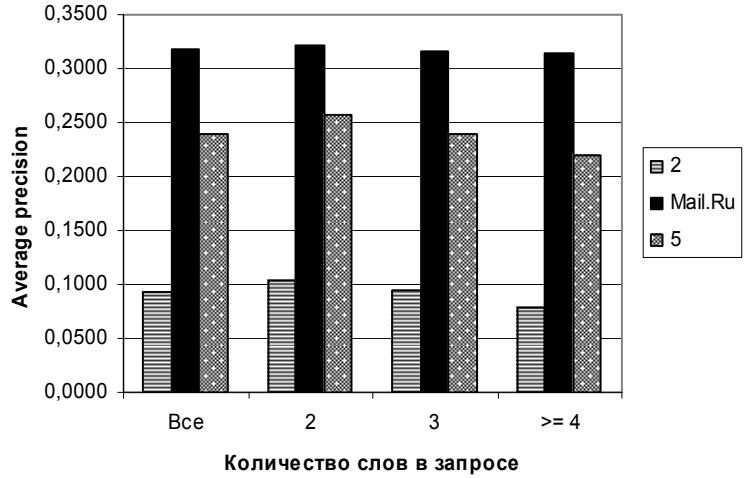


Рисунок 3. Плавное снижение AvP.

Во-вторых, мы решили исследовать зависимость AvP от размера пула. Были выбраны группы «меньше 150 документов в пуле» (11 запросов), «150-200» (20), «200-250» (24), «250-300» (11), «больше 300» (9). Предполагалось, что больший размер пула по запросу соответствует более неопределенному поведению всех участников в совокупности и, как следствие, более сложным для участников запросам. Из результатов (см. рис. 5) видно, что на неоднозначных

запросах для нашей системы оказалось характерным более высокое, в сравнении с другими участниками, качество работы, чем в среднем по всей дорожке. На «простых» запросах наша система также показала значительно лучшие результаты, чем в среднем, хотя ожидалось, что по этим запросам как раз показатели разных участников будут мало отличаться друг от друга. На «среднесложных» же запросах большинство участников, напротив, показывают более близкие результаты.

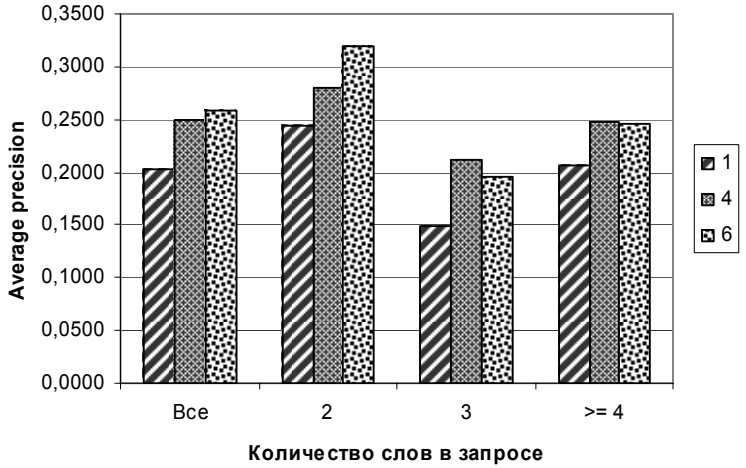


Рисунок 4. Скачок AvP между 2 и 3 словами.

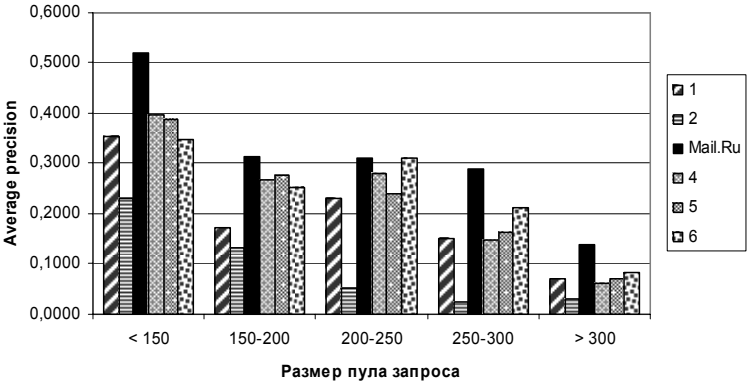


Рисунок 5. Зависимость AvP участников от размера пула запроса.

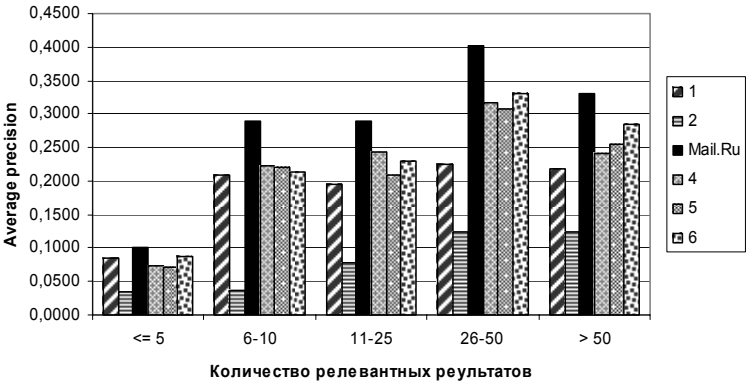


Рисунок 6. Зависимость AvP участников от количества релевантных результатов.

В-третьих, было рассмотрено разбиение запросов по количеству релевантных, по мнению оценщиков, документов. Были выделены группы «менее 5 релевантных» (6 запросов), «6-10» (10), «11-25» (20), «26-50» (23), «более 50» (16). Сравнительная картина участников получилась похожая на всех рубриках (см. рис. 6).

Абсолютные же значения показывают, что, например, у всех систем плохо получается отвечать на запросы с малым количеством релевантных документов. Это может быть вызвано несколькими причинами. Возможно, некоторые из этих запросов настолько сложны для систем, что, несмотря на наличие в коллекции релевантных запросу документов, ни одна система не выдала приемлемого их количества и, соответственно, в пул попала лишь очень небольшая их часть. Впрочем, в группе содержится только 6 запросов и это вполне может оказаться флуктуацией.

Данные исследования, безусловно, являются пробным камнем и не претендуют на абсолютную объективность, особенно ввиду малого количества запросов. Тем не менее мы считаем важным изучение устойчивости поведения систем на разных подмножествах запросов и надеемся увидеть эти или другие подобные закономерности официально изучаемыми на одном из следующих семинаров.

2.3.2 Legal-adhoc, Mixed-adhoc

Для дорожки поиска по нормативным документам (Legal-adhoc) участникам компанией «Кодекс» была предоставлена коллекция

размером около 67000 документов. Полностью аналогично дорожке веб-поиска, задача состояла в выполнении большого количества (12900) запросов на этой коллекции. Ответом системы также считалось до 100 наиболее релевантных документов в выдаче.

Для дорожки смешанного поиска необходимо было на объединенной коллекции веб- и нормативных документов выполнить объединенное множество запросов, предоставленных для дорожек Web-adhoc и Legal-adhoc. Целью являлась проверка возможности систем функционировать в сильно разнородной коллекции, так как веб- и нормативные документы значительно отличаются по своим характеристикам.

Для выполнения заданий этих двух дорожек использовалась та же поисковая система, отличие состояло только в настройке параметров для конкретной задачи.

Так, например, в нормативных документах заголовок скорее всего будет содержать слова, напрямую относящиеся к тематике документа. В то же время в веб-коллекции мусор в заголовках страниц – вполне обычное дело. Соответственно, был изменен вес для слов и пассажей, входящих в важные зоны документа.

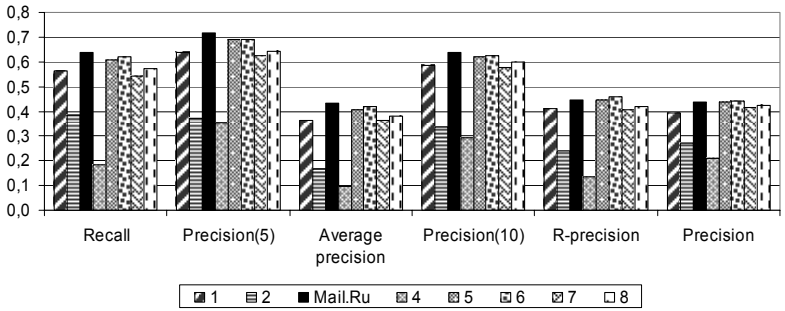


Рисунок 7. Результаты дорожки Legal-adhoc по оценке legal-pd50-all

Run	1	2	**Mail.Ru**	4	5	6	7	8
Recall	0,5644	0,3834	**0,6366**	0,1855	0,6098	0,6212	0,5416	0,5718
P(5)	0,6370	0,3728	**0,7160**	0,3556	0,6889	0,6914	0,6272	0,6444
AvP	0,3644	0,1677	**0,4339**	0,0949	0,4071	0,4178	0,3632	0,3805
P(10)	0,5852	0,3346	**0,6370**	0,2951	0,6222	0,6272	0,5753	0,5988
R-prec.	0,4105	0,2414	**0,4440**	0,1347	0,4459	0,4592	0,4071	0,4211
Precision	0,3943	0,2713	**0,4371**	0,2104	0,4379	0,4421	0,4148	0,4249

Таблица 3. Те же результаты в табличной форме

Также в правовых документах чаще встречаются сложные синтаксические структуры, в результате чего связанные по смыслу слова оказываются разделенными б**о**льшим количеством сторонних слов. Для учета этого были изменены максимально возможная длина пассажа и ряд бонусов и штрафов, начисляемых за связность слов запроса в документе. И т. п.

При ручной оценке релевантности для дорожки Legal-adhoc каждый документ получил только одну оценку, в отличие от Web-adhoc, где документы получали по 2-3 оценки. В результате для дорожки Web-adhoc было возможным построить оценки AND и OR, а для Legal-adhoc – нет. Зато в этом году документы были оценены специалистами с юридическим образованием, что должно было поднять качество оценки.

Как видно из таблицы 3 и графика TREC (рис. 8), сразу несколько систем показали практически одинаково хороший результат.

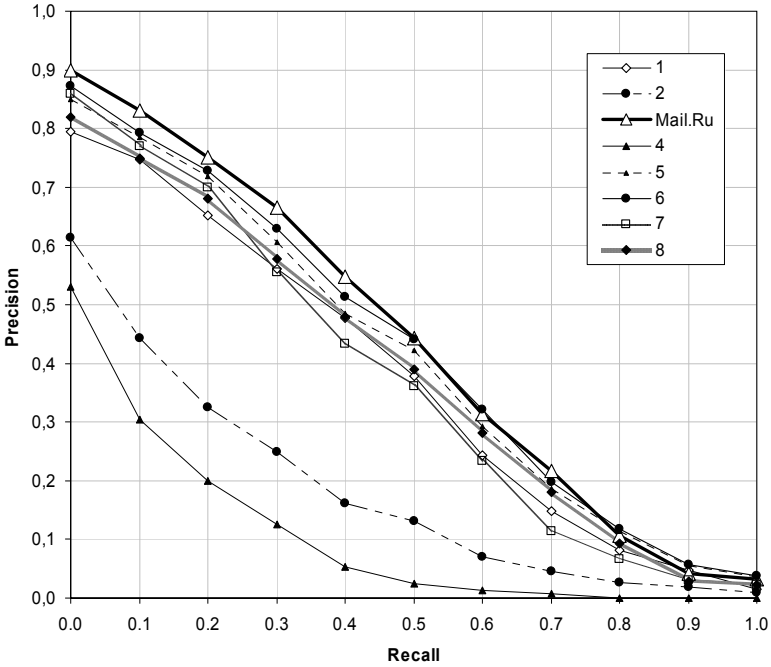


Рисунок 8. График TREC для оценки legal-adhoc-pd50-all

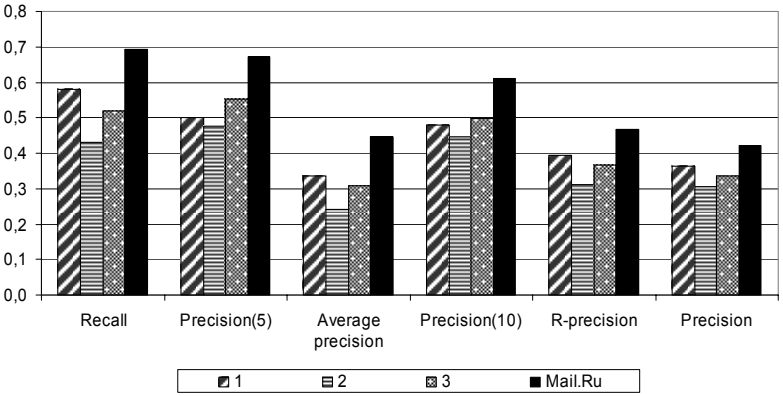


Рисунок 9. Результаты дорожки Mixed-adhoc по оценке pd50-or

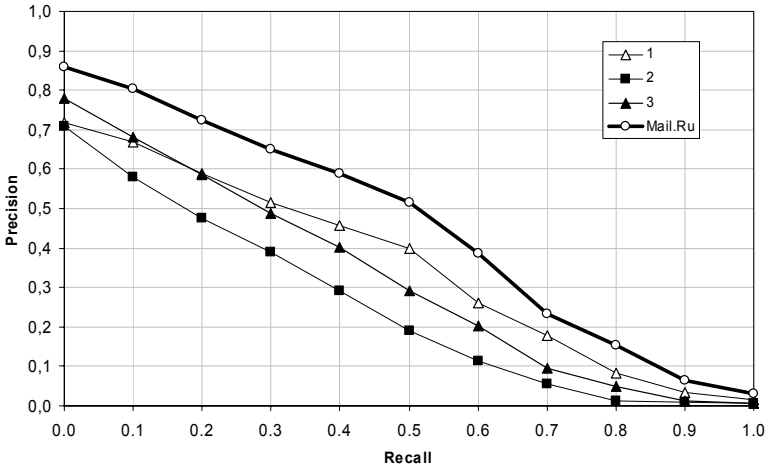


Рисунок 10. График TREC для дорожки Mixed-adhoc (pd50-or).

	все документы		pd50	
	AND	OR	AND	OR
Recall	0,8169	0,7975	0,7110	0,6932
Precision(5)	0,6014	0,6723	0,6014	0,6723
Average precision	0,4703	0,4910	0,4315	0,4471
Precision(10)	0,5295	0,6103	0,5295	0,6103
R-precision	0,4674	0,4843	0,4523	0,4674
Precision	0,2135	0,2534	0,3505	0,4202

Таблица 3. Результаты Поиск@Mail.Ru для различных оценок дорожки Mixed-adhoc.

3. Дорожки классификации

3.1 Используемые алгоритмы

Наша система классификации представляла собой решение на базе известного алгоритма SVM [7,8], хорошо зарекомендовавшего себя в задачах классификации текстов. При этом наше внимание было сосредоточено, с одной стороны, на качественной предобработке текста и выборе параметров и граничных значений при построении векторов, с другой - на экспериментах с различными характеристиками SVM алгоритма.

В частности, одной из задач, бывших предметом нашего исследования, являлся способ отсечения малоинформативных термов при построении векторов. В результате проведенных исследований мы остановились на игнорировании редких термов, исходя из их подокументной встречаемости, а также термов, встречающихся хотя бы в одном документе в проценте категорий обучающей коллекции, большем заданного.

К сожалению, ограниченность во времени при подготовке к семинару не позволила нам провести все задуманные исследования в полном объеме; некоторые важные решения принимались нами интуитивно. Тем не менее, в целом, наша система показала сравнительно неплохие результаты, в особенности при классификации нормативно-правовых документов.

3.2 Участие в дорожках

На семинаре мы представили 4 классификационных прогона:

- классификация веб-страниц (web-class) – 1 прогон;
- классификация веб-сайтов (webpage-class) – 1 прогон;
- классификация нормативных документов(legal-class) - 2 прогона.

Результаты по классификации веб-сайтов и страниц оказались весьма средними, тут есть еще над чем поработать.

Для обеих этих дорожек использовался SVM-метод с ядром RBF. Основное внимание было уделено тюнингу параметров алгоритма и качественной предобработке данных.

При группировке страниц в сайты исследовались различные весовые функции для страниц. В частности, мы изучали, возможно ли использовать «меру соответствия» страницы рубрике, полученную на выходе алгоритма SVM или его можно использовать лишь как бинарный классификатор. В итоге наших экспериментов выяснилось, что учет меры соответствия повышает значения оценочных параметров. Изучалась также возможность отфильтровывать заведомо мусорные страницы, но здесь мы сильно не продвинулись, требуется дальнейшее изучение.

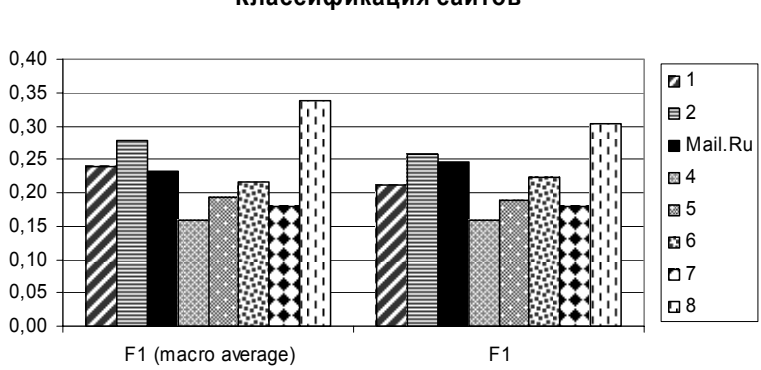


Рисунок 11. Значения мер F1 и F1-macro для участников дорожки классификации сайтов со слабыми требованиями к релевантности.

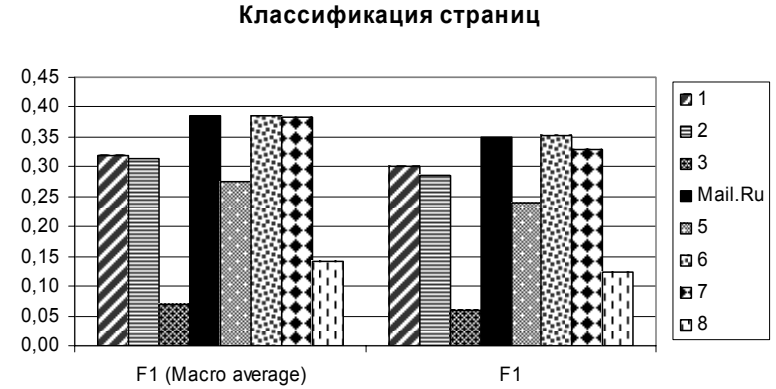
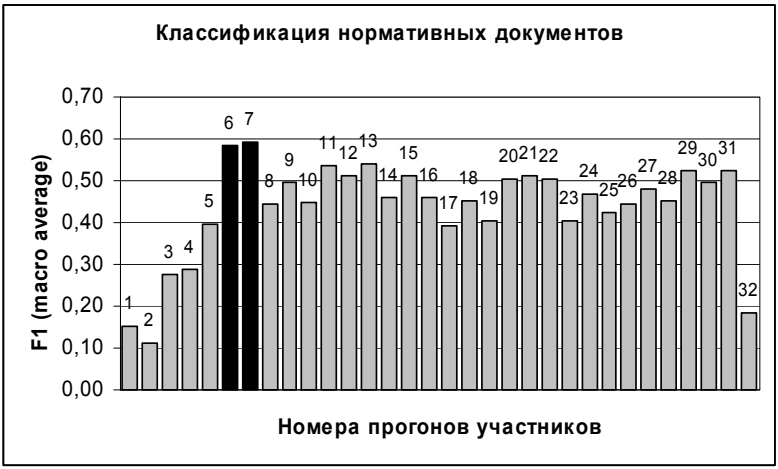


Рисунок 12. Значения мер F1 и F1-macro для участников дорожки классификации страниц со слабыми требованиями к релевантности.

Для классификации нормативных документов помимо RBF (прогон №6), использовалось также полиномиальное ядро (прогон №7). Результаты для последнего оказались немного лучше, но процесс классификации с помощью него проходил в несколько раз раза медленнее. По этой причине больше исследованым оказалось именно ядро RBF.



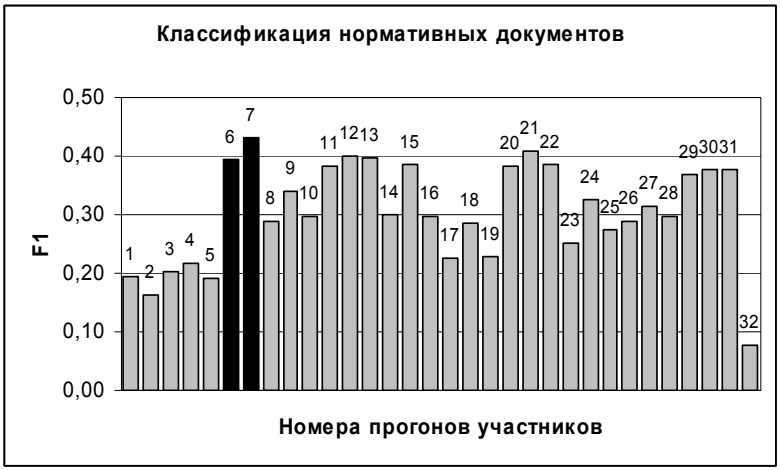


Рисунок 13. Значения мер F1 и F1-macro для участников дорожки классификации нормативных документов со слабыми требованиями к релевантности.

4. Заключение

В целом мы вполне удовлетворены результатами, которые показали наши системы на дорожках семинара. Поисковая система достигла прекрасных показателей на всех трех предложенных коллекциях. Классические алгоритмы классификации оказадись вполне применяемыми. Однако еще придется поработать как над непосредственной оптимизацией алгоритмов классификации, так и со стадиями предобработки, предварительной фильтрации документов обучающей коллекции и последующей группировки страниц в сайты.

Литература

[1] J. P. Callan. Passage-level evidence in document retrieval. In *The 17 Conference on Research and Development in Information Retrieval,* pages 302-309, Dublin, Ireland, 1994. ACM

[2] S. E. Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, and M. Gatford. Okapi at trec-3. In *TREC-3*, 1994.

[3] G. Salton, C. Buckley. Term-weighting approaches in automatic text retrieval. In *Information Processing and Management: an International Journal,* Volume 24, Issue 5, pages: 513 – 523, 1988.

[4] S. Brin, L. Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. In *Computer Networks and ISDN Systems,* Volume 30, number 1-7, pages 107-117, 1998.
[5] Под ред. И.С. Некрестьянова. Труды РОМИП'2004 *Санкт-Петербург: НИИ Химии СПбГУ,* 214 с, сентябрь 2004
[6] М.С. Агеев, Б.В. Добров, Н.В.Лукашевич, А.В. Сидоров. Экспериментальные алгоритмы поиска/классификации и сравнение с "basic line". *Труды второго российского семинара по оценке методов информационного поиска*. *Под ред. И.С. Некрестьянова* - Санкт-Петербург: НИИ Химии СПбГУ, 2004, 214 с.
[7] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. Technical Report 23, Universitat Dortmund, LS VIII, 1997
[8] C. Burges. A tutorial on support vector machines for pattern recognition. In Data Mining and Knowledge Discovery, volume 2, pages 1 – 43. Kluwer Academic Publishers, Boston, 1998

Mail.Ru at RIRES 2005

A. Fedorovsky, M. Kostin, A. Proskurin
Mail.Ru
fedorovsky@corp.mail.ru, kostin@corp.mail.ru,
proskurin@corp.mail.ru

The article presents information retrieval system Search@Mail.Ru and a set of classification algorithms developed by Mail.Ru company at RIRES. A brief description of Search@Mail.Ru internal architecture is given, emphasized at structure relevance function and influence of its parts to obtained results of passing RIRES tracks.