

Кто виноват и где собака зарыта?

Метод валидации ответов на основе неточного сравнения семантических графов в вопросно-ответной системе.

А.А.Соловьёв
МГТУ им. Н.Э. Баумана

a-soloviev@mail.ru

РОМИП 2010, 15 октября, Казань

План доклада

- Цели и задачи
- Типовая реализация
вопросно-ответной системы
- Метод валидации ответа нестрогим
сравнением семантических графов
- Эксперименты на РОМИП 2010
- Заключение

Цель и задачи

Целью работы является экспериментальное исследование методов вопросно-ответного поиска, применённых к русскому языку.

Для достижения этой цели были поставлены следующие задачи:

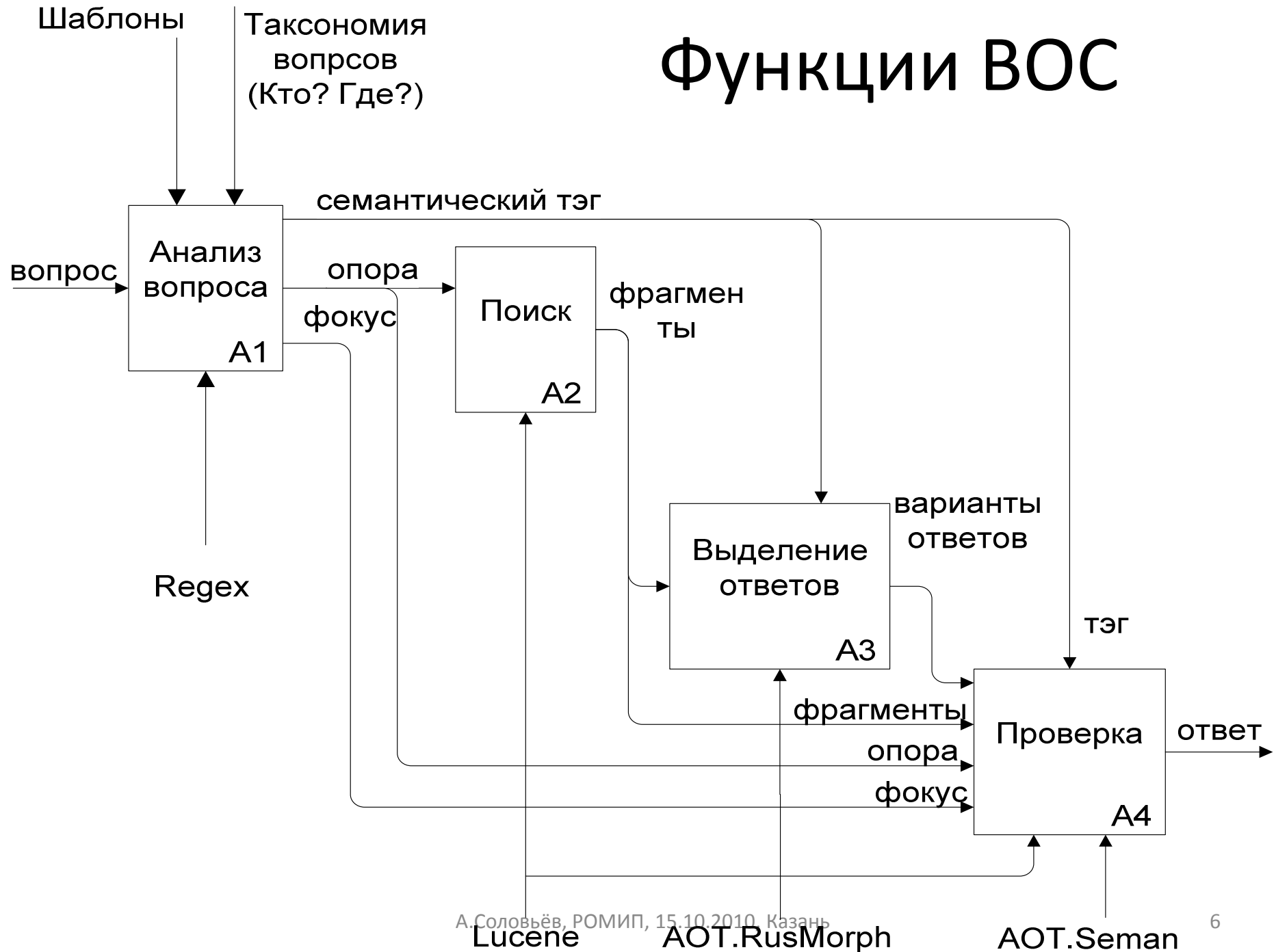
- Изучить существующие методы и системы.
- Построить экспериментальную вопросно-ответную систему.
- Найти и приспособить доступные русскоязычные компоненты компьютерной лингвистики для задачи вопросно-ответного поиска.
- Провести экспериментальную оценку качества работы системы в различных конфигурациях.

Типовая реализация вопросно-ответной системы

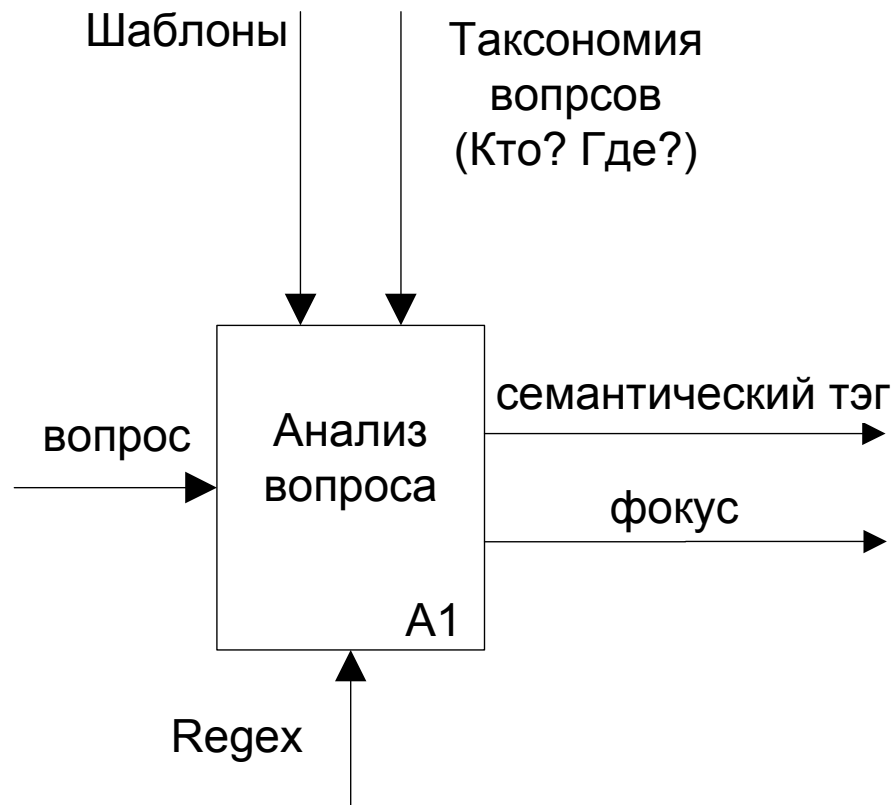
Типовая архитектура – метапоисковая система



Функции ВОС



Анализ вопроса



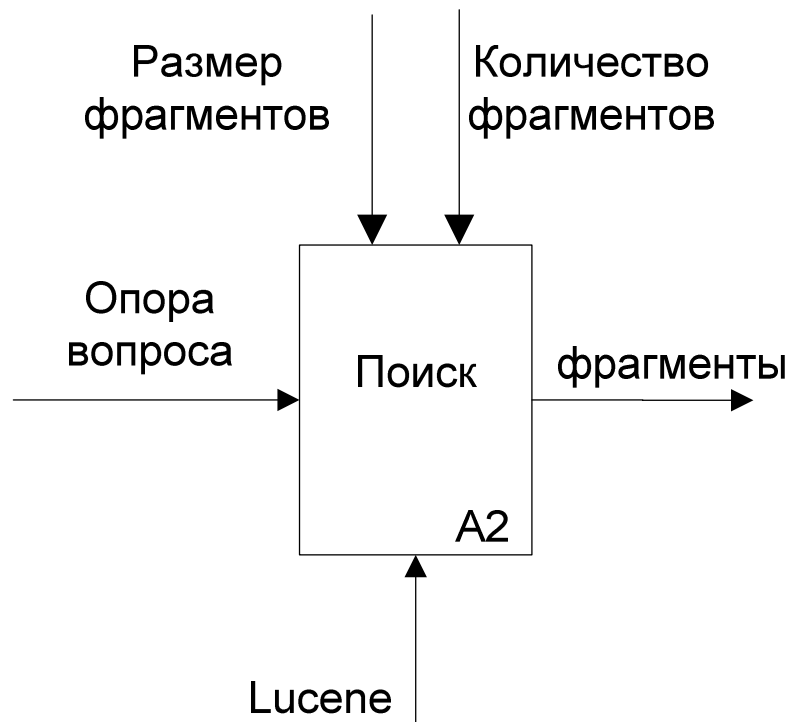
*Кто использовал
стволовые клетки?*

Тэг: PERSON

*Где собирают
мегафоны?*

Тэг: LOCATION

Поиск



Яндекс

Нашлось
60 тыс. ответов

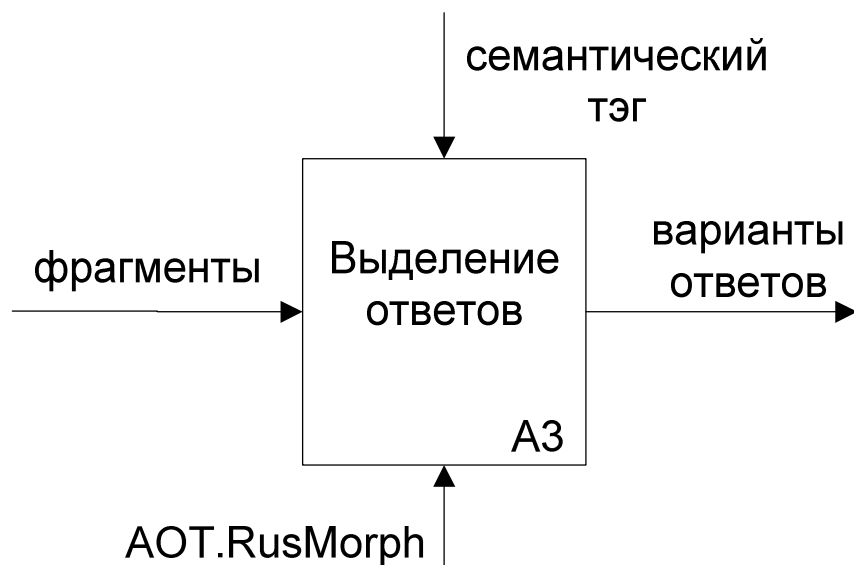
где собирают меганы?

в найденном в Москве

расширенный по

Быть может, вы искали: «[где собирают меган?](#)»

- [а где еще меганы собирают? - форум Renault Megane \(Рено Меган\) Россия](#)
Входящий в тему. Группа: Пользователи. Сообщений: 90. Регистрация: 5.8.2008. Из: Москва. Пользователь №: 8381. Заметила на улицах другой вид мегана. У него в комплектации Ауфентик, другого цвета пластик по борту (светлее, чем у меня) и на этом...
[www.megane-club.ru > Форум > Россия > ?showtopic=14116](#) [копия](#) [ещё](#)
- [Renault Club Forum > Де насправді налагоджений випуск RENAULT?](#)
В Бурсе . **где собирают** седаны **меганы** и символы. там №32. в том же городе находится заводы по сборке "каролл" и "фокусов" ... так он уверял, что с 2004-го года все **Меганы собираются** только в Турции, а покупателям, типа, лапшу вешают про французскую...
[www.renault-club.kiev.ua > В начало форума > .../index.php/t14454.html](#) [копия](#) [ещё](#)
- [За рулем](#)
Естественно, в разговоре с К. Гакомсом я не мог не затронуть тему открытия в Москве совместного предприятия с "Рено", **где будут собирать "меганы"**. Получается, вместо бельгийского "Рено" создает завод в российской столице.
[www.zr.ru/a/6644/print.html](#) [копия](#) [ещё](#)
- [ООО "Леальта" :: Просмотр темы - Вопрос владельцам Lancer](#)
Чтобы не вступать в полемику точно вечером напишу, **где Меганы собирают**. Logan 100%, Меган не намного лучше. И тем более Турция тоже не вариант для сборки авто.
[www.lealta.ru/forum/viewtopic.php...](#) [копия](#) [ещё](#)
- [Журнал "Автопилот" номер 12 за 2003г.: TEST-DRIVE:](#)
Традиционно седаны любят в России, а еще в Турции, **где "Меганы"** с багажником и будут **собирают**. Страна сборки никого огорчать не должна, хотя, наверное, и может: известно, что в Турции **собирают** не самый престижный Renault - Symbol, в то время как...
[autopilot.kommersant.ru/Issues/Auto/2003/12/...](#) [копия](#)



Выделение ответов

Где собирают меганы?

Тэг: LOCATION

[Renault Club Forum > Де насправді налагоджений випуск RENAULT?](#)

В Бурсе . где собирают седаны меганы и символы. там №32. в том же **городе** находится заводы по сборке "каролл" и "фокусов" ... так он уверял, что с 2004-го года все Меганы собираются только **в Турции**, а покупателям, типа, лапшу вешают про французскую...

www.renault-club.kiev.ua

[Журнал "Автопилот" номер 12 за 2003г.: TEST-DRIVE:](#)

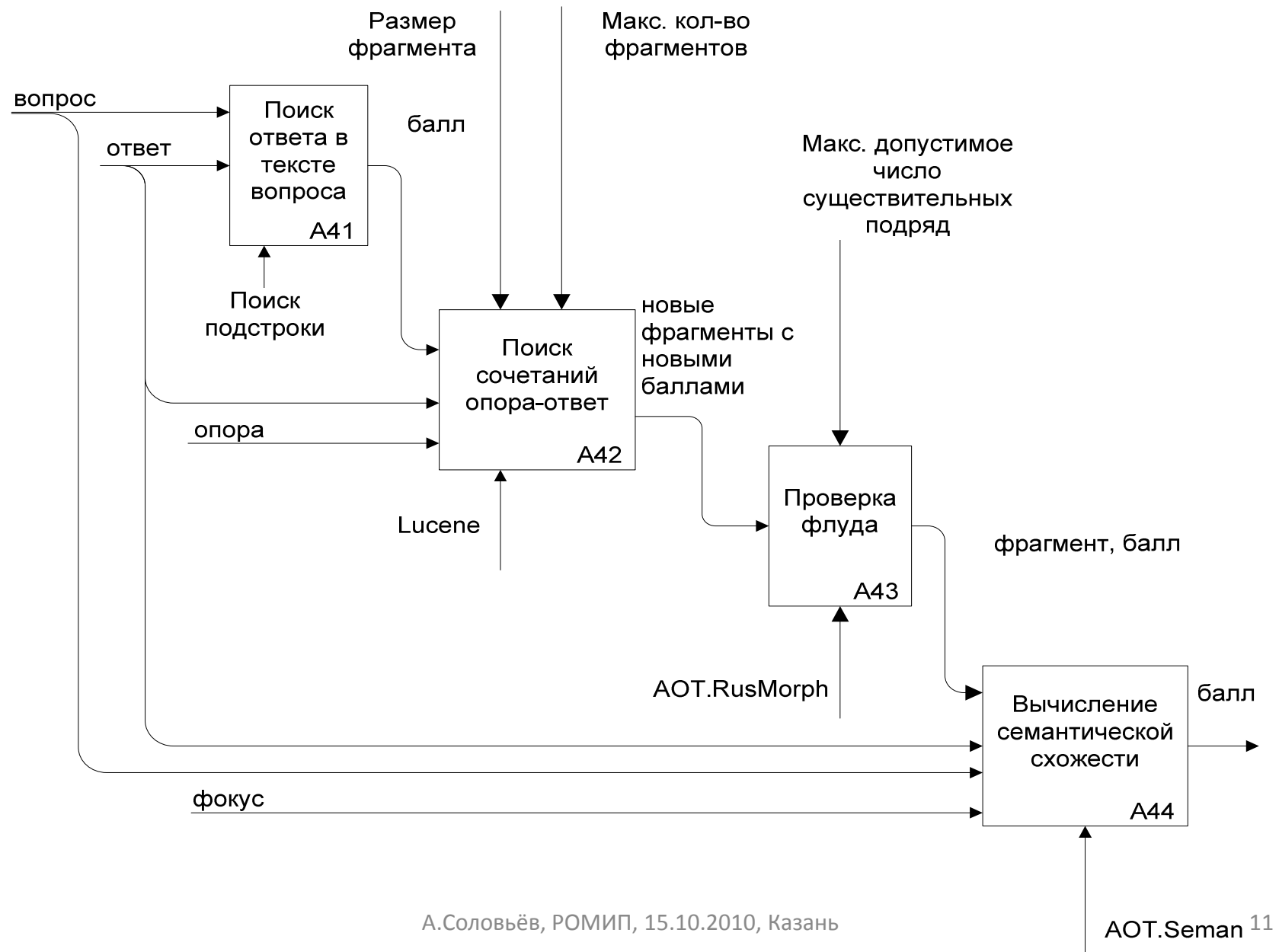
Традиционно седаны любят **в России**, а еще **в Турции**, где "Меганы" с багажником и будут собирать. **Страна** сборки никого огорчать не должна, хотя, наверное, и может: известно, что **в Турции** собирают не самый престижный Renault - Symbol, в то время как...

autopilot.kommersant.ru/Issues/Auto/2003/12/

Валидация ответов: функциональная схема типового фильтра

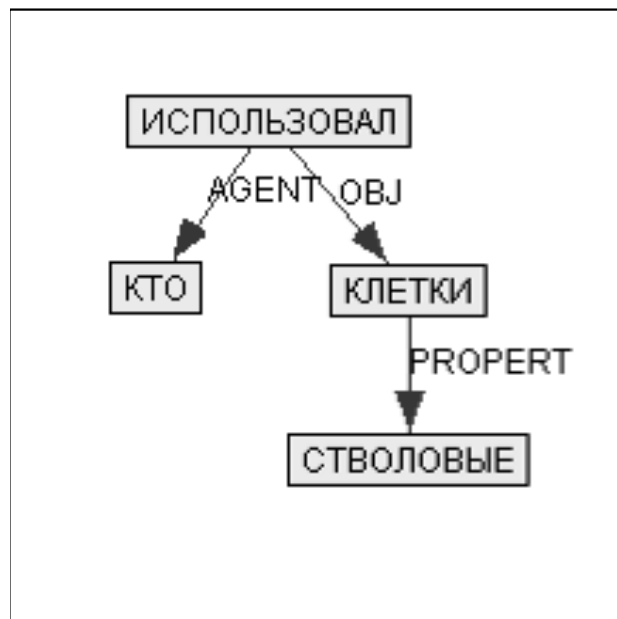
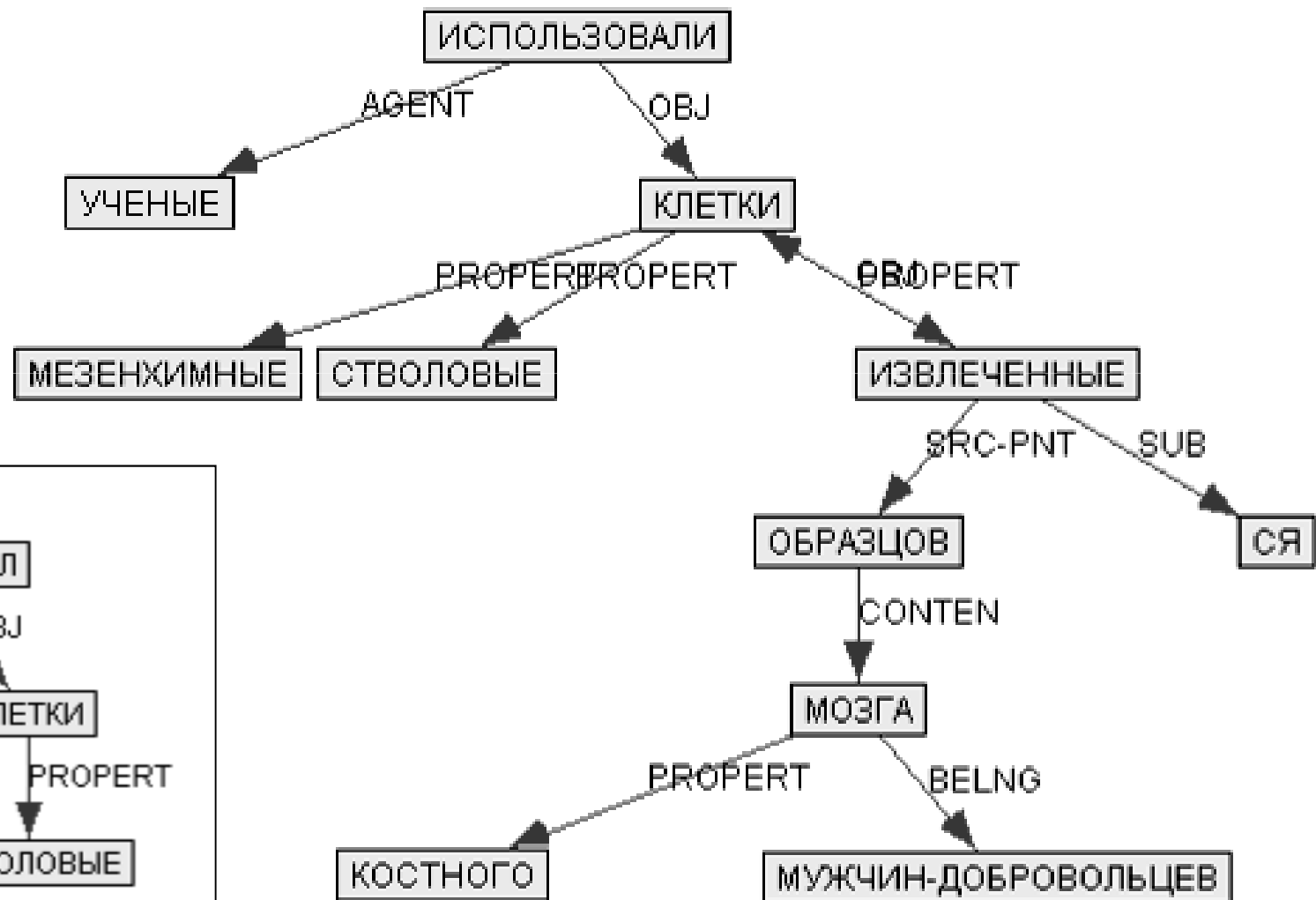


Валидация ответов: каскадная проверка

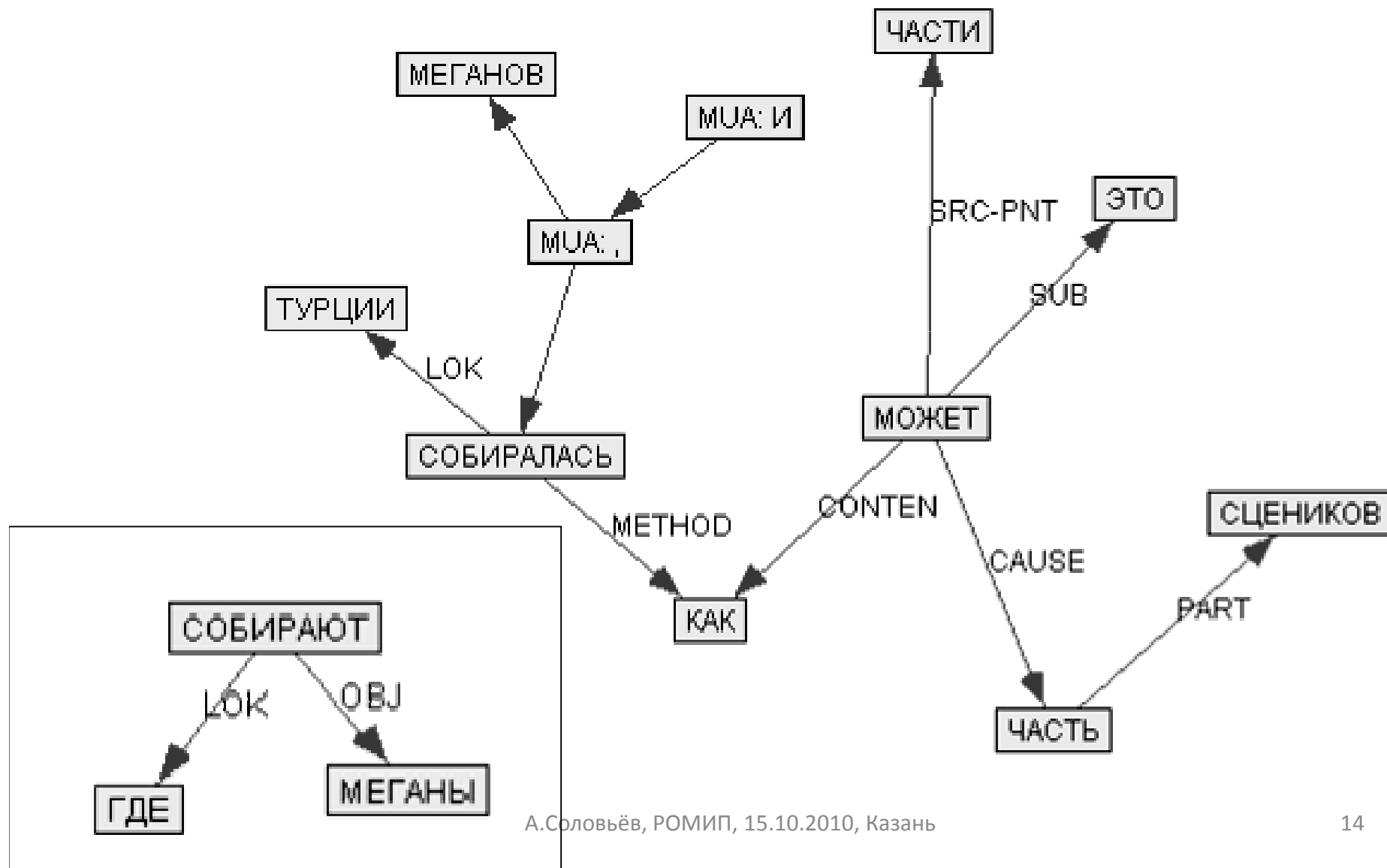


Метод валидации ответа
нестрогим сравнением
семантических графов

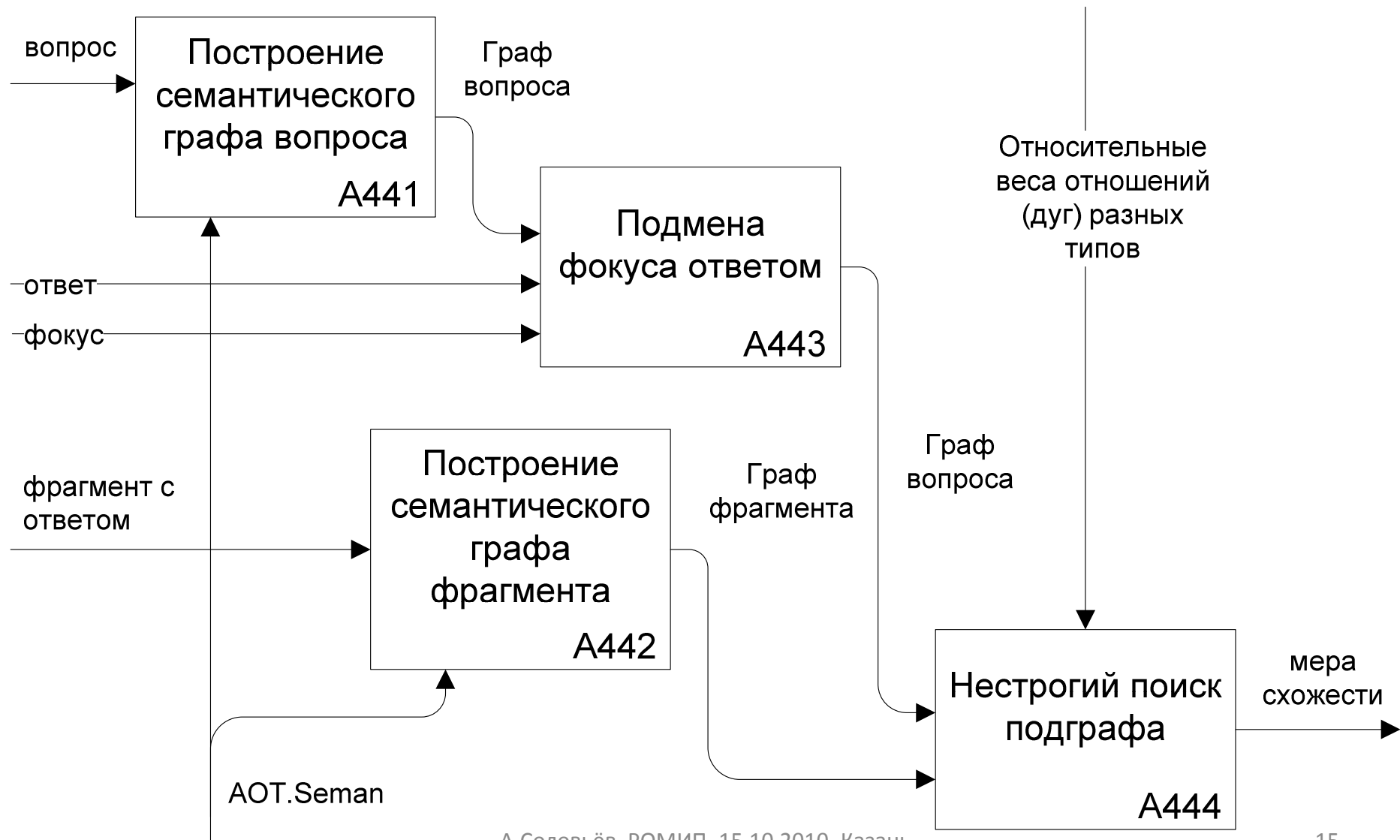
Метод валидации ответа на основе неточного сравнения семантических графов



Метод валидации ответа на основе неточного сравнения семантических графов



Метод валидации ответа на основе неточного сравнения семантических графов



Эксперименты на РОМИП 2010

Общие результаты QA на РОМИП2010

- Участников: 2
- Прогонов: 4
- Коллекция: VU.web 2007 ~1.5 млн. страниц
- Вопросов сдано: ~10 тыс.
- Вопросов отобрано организаторами: 246
- Правильных ответов: 14
- Вопросов, потенциально имеющих ответ в коллекции: 60

Правильные ответы (1-7 из 14)

Вопрос	Фрагмент с ответом
где снимали фильм сталкер?	Солоницын, Николай Гринько, Алиса Фрейдлих, Файме Юрна, Наташа Абр Производство: Россия Год: 1979. Продолжительность: 2.36. Краткое описание: Андрей Тарковский сказал о фильме "Сталкер" так: "Я готовился к фильму всю жизнь, снимал его два года". В этой ленте, снятой по мотивам повести братьев
чем питаются мальки рыб?	Первые дни новорожденные питаются микроводорослями, через 2 3 суток их потихоньку начинают подкармливать разведенными пекарскими дрожжами, живой «пылью»
что такое акдс?	АКДС адсорбированная коклюшно дифтерийно столбнячная вакцина. Адсорбированная коклюшно дифтерийно столбнячная вакцина (АКДС). Прогрессирующие заболевания
"кто входит в совет федерации?"	С 1995 г. в Совет Федерации входили губернаторы и спикеры законодательных собраний субъектов РФ. Депутат ГД Алексей Розуван и член Совета Федерации Андрей Ищук подго
где отдыхали летом?	Регистрация: 27.08.2006 Адрес: Сілічы Сообщения: 1,731 алушта (я там летом отдыхал) Падонак Посмотреть профиль Отправить личное сообщение для Падонак Найти еще сообщения
23 февраля, выходной день?	23 февраля Г не красный день календаря,. Не красный, знать, не выходной,. Не красный, нет, он золотой. 23 ФЕВРАЛЯ ОТ А ДО Я! День рождения.
кто построил парфенон???	— Аристотель, Диоген, Фидий, который построил Парфенон в Древней Греции, а потом был жестоко наказан за то, что изобразил там где – то в углу свое собственное лицо: т

Правильные ответы (7-14 из 14)

Вопрос	Фрагмент с ответом
биография чингиз айтматов умер где?	Чингиз Торекулович Айтматов. Чингиз Торекулович Айтматов родился в 1928 г. в кишлаке Шекер в Киргизии. Наиболее известные работы Айтматова:
где сейчас играет футболист марат измайлов?	Полузащитники — Динияр Билялетдинов, Марат Измайлов, Дмитрий Лоськов (все — «Локомотив» Москва), Евгений Алдонин (ЦСКА), Владислав Радимов («Зенит» Санкт-Петербург), Игорь Семшов («Динамо»); Нападающие — Андрей Аршавин, Александр Кержаков (оба — «Зенит»),
март по гороскопу кто ?	Хотим ребенка » Планируем характер малыша » Влияние планет на дату рождения » Малыш ОВЕН 21 МАРТА - 20 АПРЕЛЯ Пока папа и мама младенца, родившегося между 21 марта и 20 апреля, принимают гостей, новорожденный овен заходится от крика в
водитель-экспедитор кто это?	Экспедитор обязан контролировать силами водителя процесс погрузки (выгрузки), включая пересчет грузовых мест, внешнее состояние упаковки, порядок погрузки, крепл
2. в какой стране изобрели панамы?	116 лет с 1337 по 1453 год.2) Панамы изобрели в Эквадоре .3) Октябрьская революция отмечается в ноябре.4) Имя короля Георга - Альберт. Он поменял его в 1936 году.5) Название Канарских островов происходит от тюленя.
кто автор знаменитой статуи дискобол?	Знаменитая статуя скульптора Мирона изображает атлета, готовящегося к метанию диска. Знаменитая статуя скульптора Поликлета (датированная второй половиной 5 в. до
15 апреля кто по гороскопу?	Телевидение Энциклопедия / Роды / Хотите верьте... Информация для авторов ЗДЕСЬ Психотип " Овен " Гороскоп для малышей родившихся с 21 марта до 20 апреля Психотип " Телец " Гороскоп для малышей родившихся с 21 апреля до 20 мая Психотип

Категории ответов системы

Прогон \ Эталон	Ответ на вопрос есть в коллекции	Правильного ответа на вопрос в коллекции нет
Система дала хотя бы один правильный ответ на вопрос	<i>a</i>	0
Система дала один или несколько ответов на вопрос, но все неправильные	<i>b</i>	<i>c</i>
Система не дала ни одного ответа на вопрос	<i>d</i>	<i>e</i>

$$E = \frac{b + c + d}{a + b + c + d + e}$$

$$R = \frac{a}{a + b + d}$$

Конфигурации трёх прогонов

- **Базовый прогон *lucene***

В базовой конфигурации система возвращала первые три распознанные именованные сущности в поисковой выдаче, соответствующие типу запроса. Единственный фильтр – ответ не должен содержаться в вопросе.

- **Прогон без семантической валидации *naive***

Во второй конфигурации были включены три фильтра:

- *Поиск ответа в тексте вопроса*

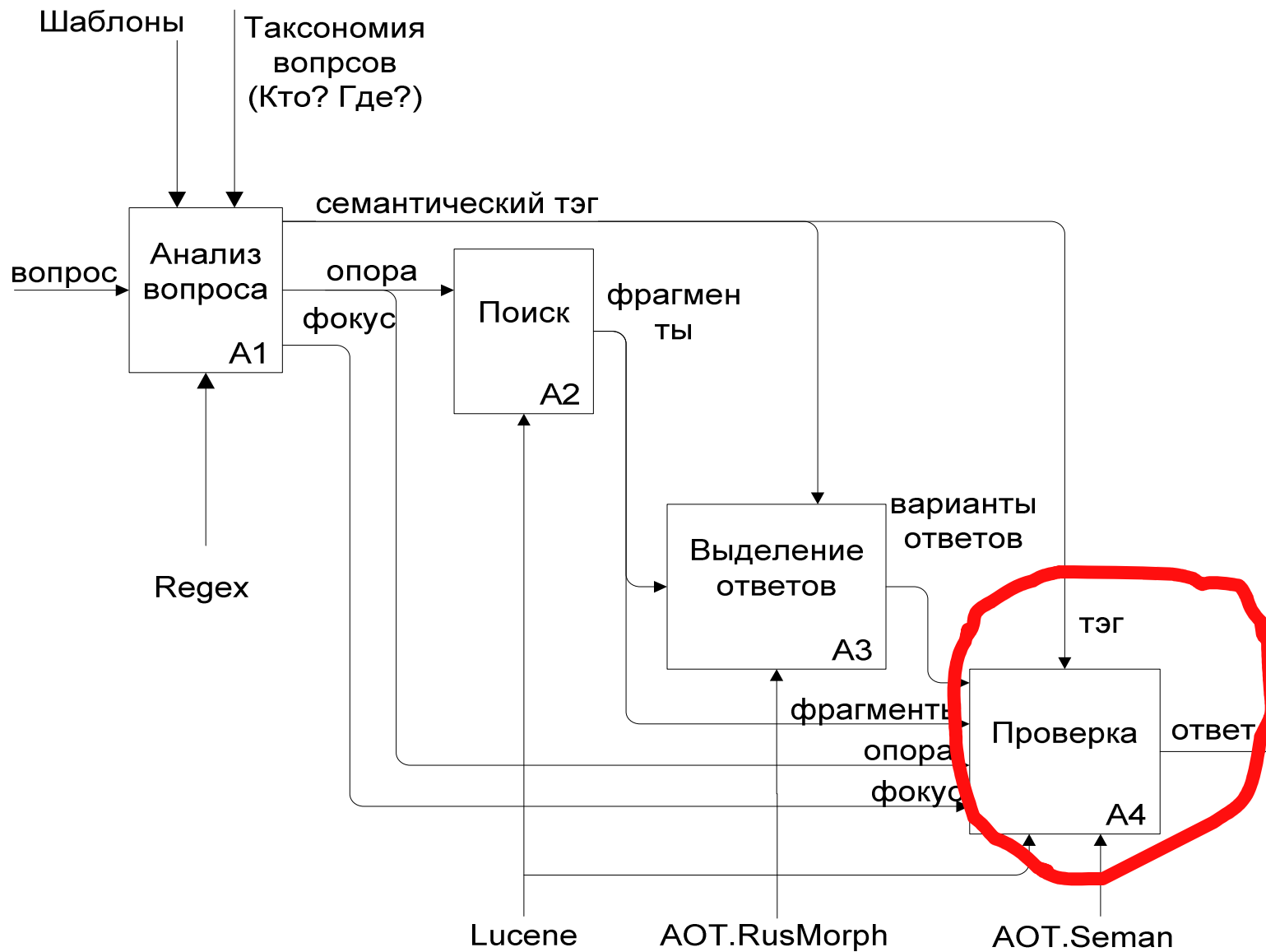
- *Поиск сочетаний опора-ответ в коллекции*

- *Проверка флуда*

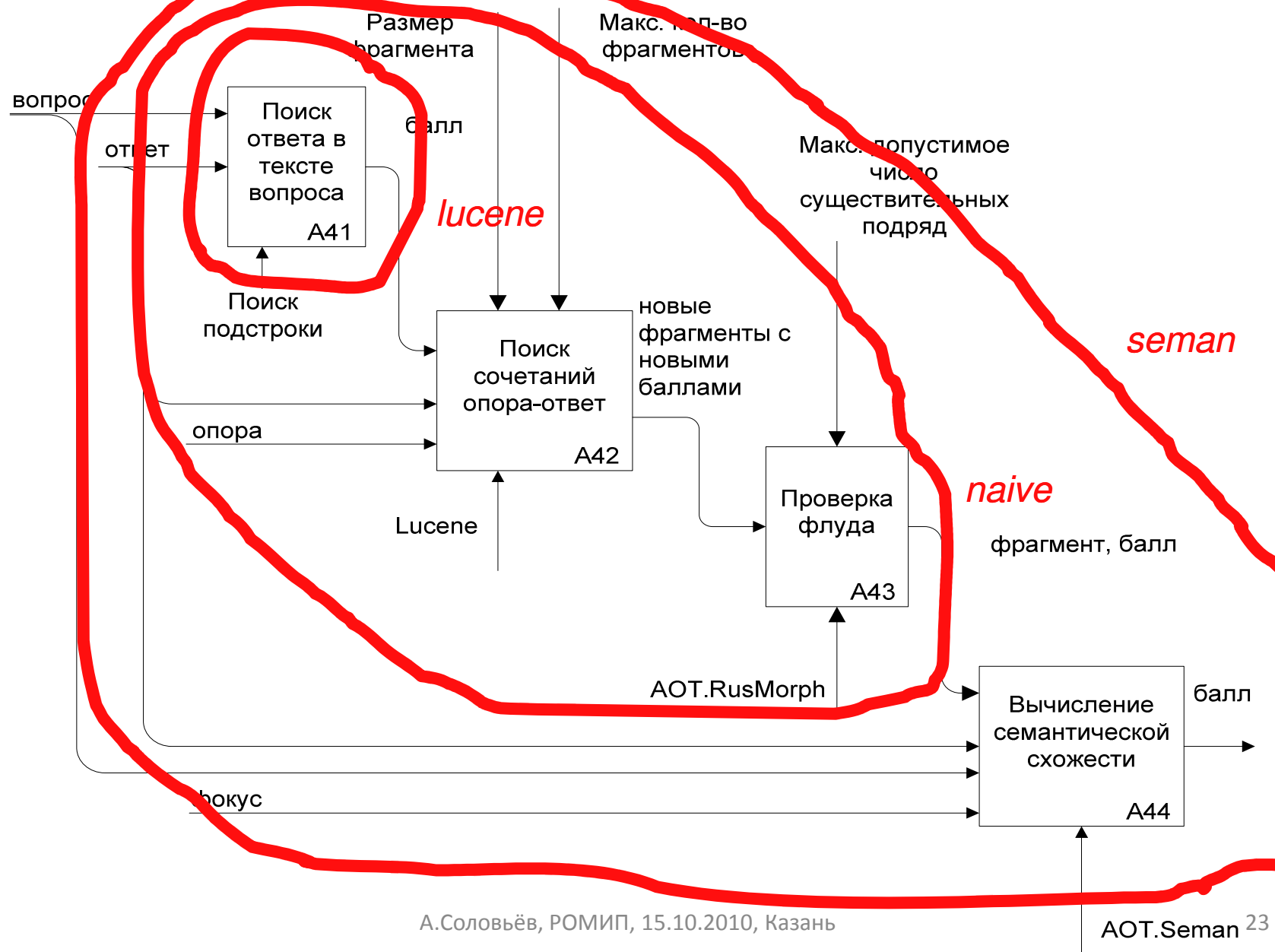
- **Прогон с семантической валидацией *setan***

Третий прогон был выполнен с полным набором фильтров.

Конфигурации трёх прогонов



Конфигурации трёх прогонов



Метрики для результатов всех прогонов

прогон/метрика	R - на сколько запросов был подан хоть один вариант ответа	Ag - число запросов у которых есть хотя бы один ответ с оценкой good	Agp1 - число запросов у которых есть хотя бы один ответ с оценкой good, long или partial	QPA - На сколько запросов из QueriesWithPotentialAnswer* была сделана попытка дать ответ	Полнота покрытия мн-ва QueriesWithPotentialAnswer* ответами с оценкой good	Полнота покрытия мн-ва QueriesWithPotentialAnswer* ответами с оценками good, long или partial	Ошибка, вычисленная по отношению к общему множеству запросов, учитывающая оценку good	Ошибка, вычисленная по отношению к общему множеству запросов, учитывающая оценки good, long и partial	Ошибка, вычисленная по отношению к мн-ву QueriesWithPotentialAnswer, учитывающая оценку good
myrtle-lucene	152	5	9	24	8,3%	15,0%	59,8%	58,1%	79,2%
myrtle-naive	149	5	9	22	8,3%	15,0%	58,5%	56,9%	77,3%
myrtle-seman	68	3	6	18	5,0%	10,0%	26,4%	25,2%	83,3%
xxx-4	246	8	39	33	13,3%	65,0%	96,7%	84,1%	75,8%
noanswer	0	0	0	0	0,0%	0,0%	0,0%	0,0%	
ideal	14	14	14	14	23,3%	23,3%	0,0%	0,0%	0,0%
всего	246	14	47	60					

*QueriesWithPotentialAnswer - запросы, у которых есть хотя бы один приблизительный ответ или документ помечен как good в общем множестве результатов всех прогонов.

Заключение

- Удалось разработать исследовательскую вопросно-ответную систему, которая отвечает на два класса вопросов: кто и где?
- Были сделаны 3 прогона в разных конфигурациях из 10 тыс. реальных вопросов пользователей на коллекции из 1.5 млн. документов.
- Применение семантического фильтра позволило уменьшить уровень ошибок с 60% до 26%, при ухудшении полноты с 8% до 5%.
- На основе полученной таблицы релевантности можно организовать повторяемый эксперимент по примеру CLEF AVE – Answer Validation Exercise.

Rodrigo, Á., Peñas, A., and Verdejo, F. Overview of the answer validation exercise 2008. In Proceedings of the 9th CLEF Conference (Aarhus, Denmark, September 17 - 19, 2008). LNCS. Springer, 296-313

Спасибо за внимание.
Вопросы?

А.А.Соловьёв
МГТУ им. Н.Э. Баумана

a-soloviev@mail.ru

РОМИП 2010, 15 октября, Казань