

# **Кто виноват и где собака зарыта? Метод валидации ответов на основе неточного сравнения семантических графов в вопросно- ответной системе**

© Александр Соловьев

МГТУ им. Н.Э. Баумана  
a-soloviev@mail.ru

## **Аннотация**

Обсуждаются эксперименты на вопросно-ответной дорожке семинара РОМИП'2010. Исследовательская система ищет в полнотекстовой коллекции краткие ответы на вопросы типа «кто?» и «где?». В основе метода лежит сравнение графов семантических связей вопроса и фрагмента текста, содержащего ответ. Автор следует типовой архитектуре вопросно-ответной системы, надстроенной поверх обычной полнотекстовой поисковой системы. Экспериментально показано, что применение предложенного семантического фильтра позволило уменьшить уровень ошибок с 60% до 26%, при ухудшении полноты выдачи с 8% до 5%.

## **Введение**

Вопросно-ответные системы – это класс информационно-поисковых систем, позволяющих использовать в качестве поисковых запросов вопросы на естественном языке и получать в результатах не просто список документов, содержащих ответ, а сам ответ – сжатый и лаконичный. Источником информации для таких систем обычно служит корпус текстов на естественном языке, например общедоступные страницы Интернет. Вопросно-ответные системы применяют методы компьютерной лингвистики (*англ.: NLP - Natural*

*Language Processing*) на каждом этапе работе и существенно зависят от качества их реализации.

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

Для достижения этой цели были поставлены следующие задачи:

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

В данной статье обсуждаются вопросы построения вопросно-ответной системы и эксперименты на вопросно-ответной дорожке семинара РОМИП'2010.

## **1. Архитектура системы**

За основу была взята типовая архитектура вопросно-ответной системы, используемая многими исследователями [1][2][4]. Система надстраивается над существующей системой поиска по ключевым словам.

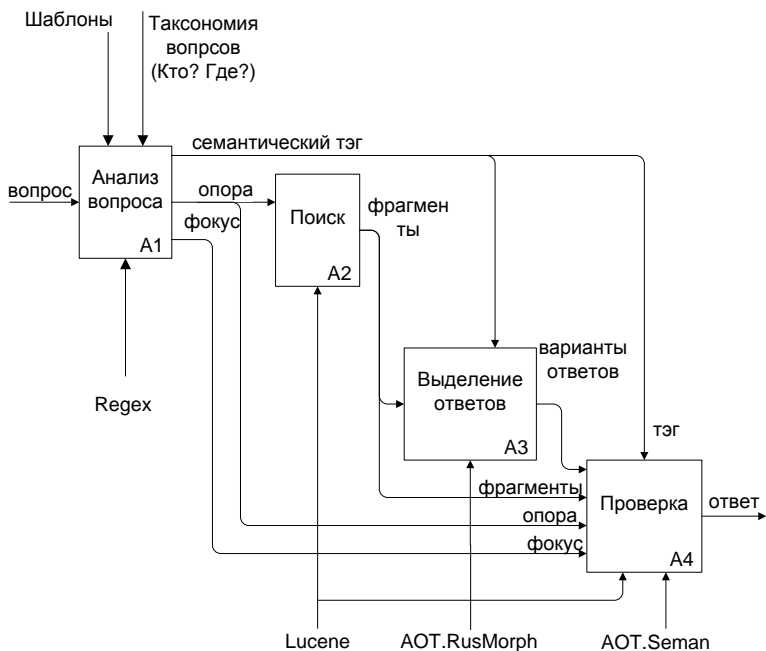
### **2.1 Функции системы**

Система соответствует разбиению задачи на подзадачи на Рис. 1. Сначала определяется тип вопроса: сейчас система распознаёт вопросы кто (о людях) и где (о топонимах). Также текст вопроса разбивается на т.н. фокус и опору. Далее на основании Опоры строится поисковый запрос Lucene [7], и выполняется полнотекстовый поиск. Результаты - фрагменты (англ.: snippet) – передаются в модуль выделения ответов.

Модуль выделения ответов проставляет семантические тэги PERSON и LOCATION используя библиотеку АОР [10]. Четвёрки <ответ,фокус,опора,фрагмент> передаются в модуль проверки вопросов.

Модуль проверки вопросов строит семантические графы для вопроса и фрагментов текста, сопоставляет их, и рассчитывает относительные веса для каждого варианта ответа. Модуль использует ещё некоторые проверки, не связанные с семантикой.

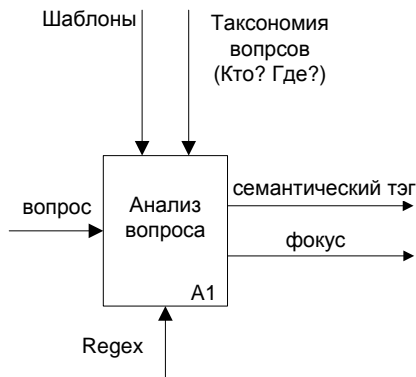
Далее рассмотрим каждую подзадачу подробнее.



**Рис. 1** Функциональная диаграмма процесса поиска ответа

## 2.2 Анализ вопросов

Анализ вопроса – это процесс простановки семантического тэга и выделения фокуса [1][12], см. Рис. 2.



**Рис. 2** Анализ вопроса

**Фокус вопроса** (англ.: question focus) – это такие сведения, содержащиеся в вопросе, которые несут в себе информацию об ожиданиях пользователя от информации в ответе [2].

**Опора вопроса** (англ.: question support) – это оставшаяся часть вопроса (после «вычета» фокуса), которая несёт в себе информацию, поддерживающую выбор конкретного ответа.

**Семантический тэг ответа** (англ.: answer tag, answer type) – класс запрашиваемой пользователем информации согласно некоторой ранее заданной таксономии.

Система использует вручную подготовленные регулярные выражения для определения одного из двух известных ей тэгов: PERSON и LOCATION. Эти же выражения используются для извлечения фокуса. Примеры используемых регулярных выражений в нотации библиотеки .NET [9]:

- Person:

```
\b(?<focus>(кто|кого))\b
```

- Location:

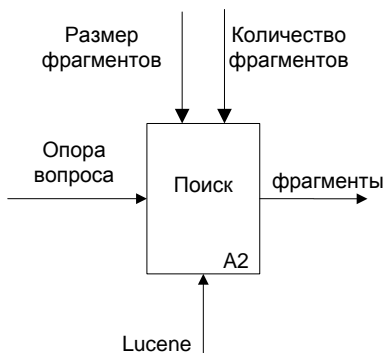
```
\b(?<focus>где(?:!+(скачать|купить)))\b
```

```
(?<focus>(в\s+)?(каком|какой|какую|какие)\s+(город\w*|стран\w*))
```

Заметим, что для вопросов кто и где фокус вырождается в вопросительное слово, поэтому не так важен.

### 2.3 Поиск фрагментов

На этапе поиска система использует *опору вопроса* в качестве поискового запроса к системе Lucene (Рис. 3).



**Рис. 3 Поиск**

В работе [7] авторы рекомендуют использовать булевский режим поиска, но в данной реализации использовался простой ранжированный поиск по ключевым словам – все слова опоры

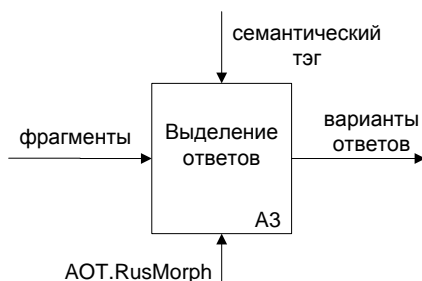
использовались в качестве ключевых слов запроса. Первые 20 фрагментов (англ.: *snippet*) поисковой выдачи будут переданы в модуль выделения ответов.

Использовалась следующая конфигурация Lucene:

- Lucene.NET 2.9.2
- Стеммер Портера (RussianStemmer из пакета Snowball)
- Разбиение на фрагменты постоянной длины – 300 символов (SimpleFragmenter из Highlighter.NET)

## 2.4 Выделение ответов

Задача выделения ответов заключается в поиске *именованных сущностей* (англ. *Named Entities*) во фрагментах, используя библиотеку AOT (Рис. 4).



**Рис. 4 Выделение ответов**

Морфологический анализ в библиотеке AOT способна проставлять грамеммы *лок*, *имя*, *фам*, *отч*, которые очевидным образом соответствуют семантическим тэгам PERSON и LOCATION. Например, в случае вопроса «кто» процесс выделения ответов найдёт все имена во фрагментах. Эти ответы-кандидаты будут переданы в модуль проверки.

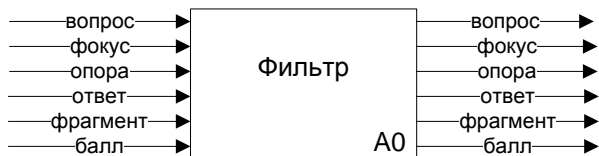
Заметим, что данная реализация плохо покрывает иностранные имена и фамилии.

## 2.5 Валидация ответов

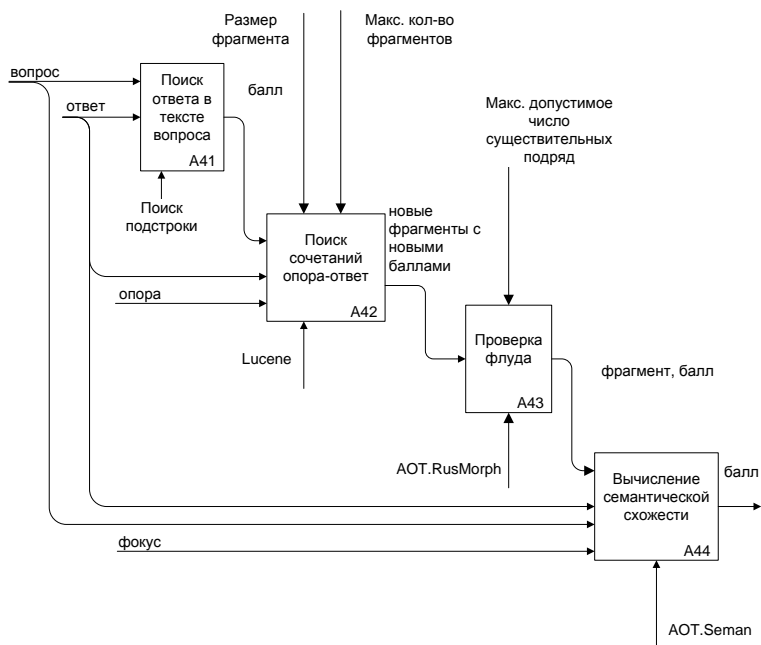
Проверка ответов заключается в фильтрации некорректных кандидатов-ответов и ранжировании оставшихся ответов. Этот этап соответствует задаче Answer Validation Exercise – CLEF AVE [3].

Проверка ответа состоит из нескольких ранжирующих фильтров, каждый из которых выставляет свой балл (Рис. 5). Фильтры соединены в каскад: выход предыдущего фильтра является входом следующего. Фильтр может выставить отрицательный или нулевой

балл, чтобы забраковать кортеж <ответ,фрагмент,балл>. Фильтр также сожжет использовать балл, выставленный предыдущим фильтром, при вычислении своего балла, т.е. корректировать балл. Реализация фильтров в системе работает со списками кортежей на входе и выходе. Такая организация фильтров используется в системе OpenEphuga [5].



**Рис. 5** Функциональная схема типового фильтра



**Рис. 6** Каскадная проверка ответа

Рассмотрим каскад из четырёх фильтров, реализованных в системе (Рис. 6):

- A41. Ответ не должен встречаться в вопросе. Например, вопрос: Где Эльбрус? Ответ: ЭЛЬБРУС. Фильтр удаляет кортежи с такими ответами.
- A42. Для каждого ответа фильтр ищет новые фрагменты по запросу <ответ> <опора>. Балл выставляется на основании числа найденных фрагментов: логарифм от этого числа. В работе [6] отмечено, что такой критерий действительно полезен.
- A43. Этот фильтр отсекает синтаксически некорректные предложения, чтобы не проводить дорогостоящий семантический анализ на следующем этапе. Выполняется морфологический анализ фрагментов с помощью AOT.RusMorph. Фрагменты с длинными цепочками существительных удаляются. Например, длинный список стран на сайте турагентства.
- A44. Вычисление меры схожести семантических графов вопроса и фрагмента, содержащего ответ (нестрогий изоморфизм подграфов).
- Во время прогонов для РОМИП [11], из ранжированного списка ответов выбирались первые 3 ответа.
- Далее рассмотрим подробнее функциональный блок A44 «Вычисление семантической схожести».

## 2. Метод валидации ответа нестрогим сравнением семантических графов

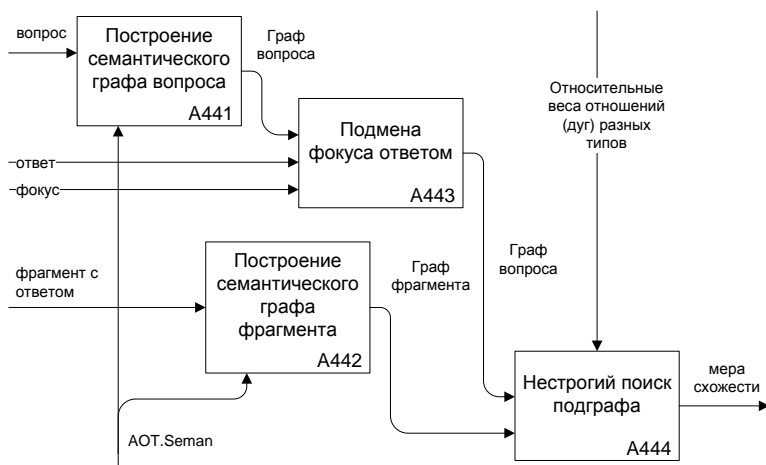
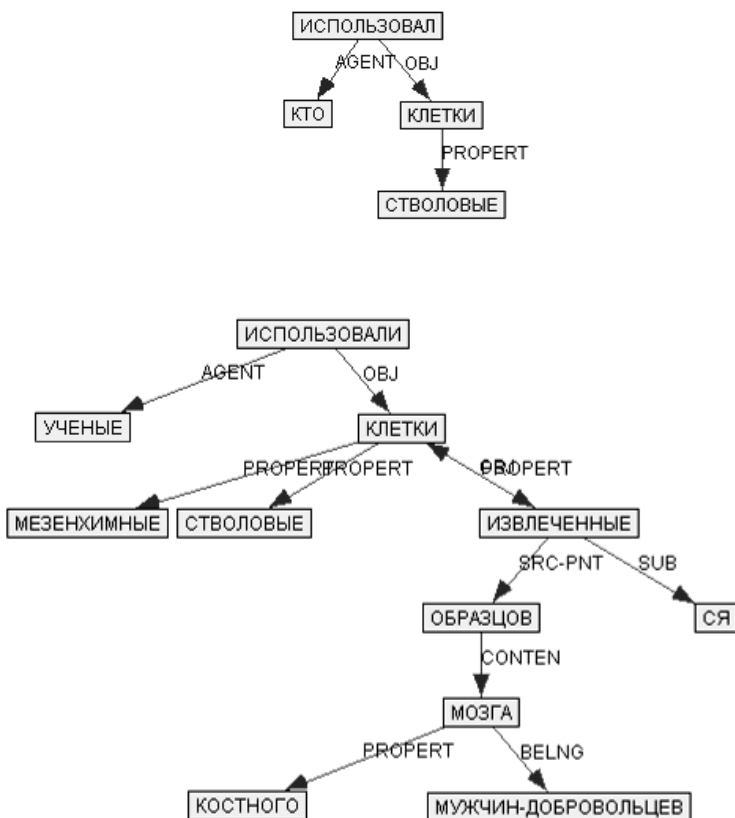


Рис. 7 Вычисление семантической схожести

Предложенный метод основан на вычислении схожести семантических графов вопроса и фрагмента, содержащего ответ. Для вопроса и фрагмента строятся семантические графы, с использованием библиотеки АОТ [10].

Рассмотрим пример семантического графа для вопроса *«кто использовал стволовые клетки?»* и фрагмента из документа 419883 *«Ученые использовали мезенхимные стволовые клетки, извлеченные из образцов костного мозга мужчин-добровольцев.»* (Рис. 8). Граф построен библиотекой АОТ.Seman.

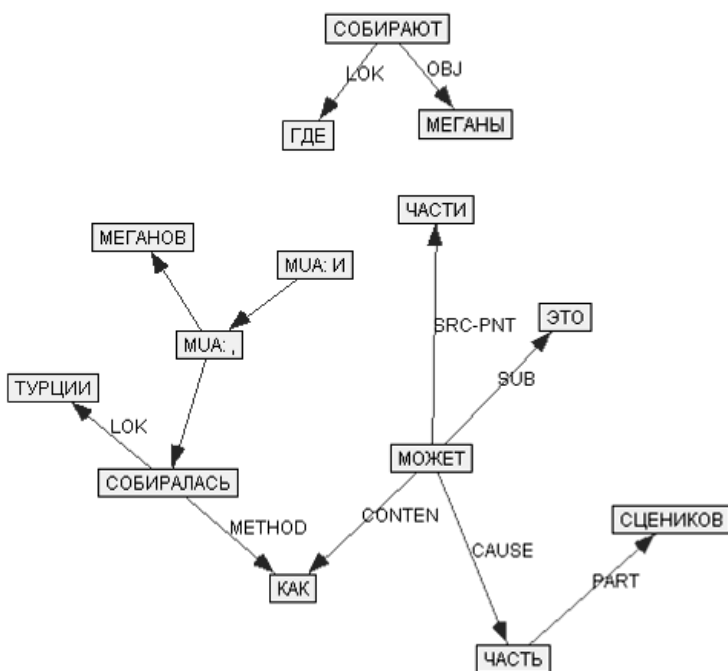


**Рис. 8 Семантические графы для вопроса «кто использовал стволовые клетки?» и фрагмента с ответом**



В основе метода лежит интуиция, что если у простого вопроса «кто?» или «где?» заменить вопросительное слово (фокус) кратким ответом, мы получим семантически верное утверждение. Мы не рассматриваем проблему синтаксической корректности полученного предложения. На Рис. 8 подграф УЧЕННЫЕ-ИСПОЛЬЗОВАЛИ-КЛЕТКИ-СТВОЛОВЫЕ во фрагменте очевидным образом соответствует графу вопроса КТО-ИСПОЛЬЗОВАЛИ-КЛЕТКИ-СТВОЛОВЫЕ, если заменить КТО на УЧЕННЫЕ. Любой строгий алгоритм поиска изоморфизма подграфов обнаружит это равенство подграфов.

Однако, более часты случаи с менее строгим совпадением подграфов. Например, вопрос nqa2009\_856: «где собирают меганы?» и фрагмент из документа 477114: «Может это от части потому, что часть Сцеников, как и Меганов, собиралась в Турции» (Рис. 9).



**Рис. 9** Семантические графы для вопроса «где собирают меганы?» и фрагмента текста с ответом.

Здесь присутствуют узлы-связки однородных членов. Стоит заметить, что дерево фрагмента в данном примере также содержит ошибку: алгоритм неправильно обработал оборот «как и».

Алгоритм вычисления меры схожести подграфов выглядит следующим образом:

1. Найти вершину с фокусом в вопросе.
2. Найти вершину с ответом во фрагменте.
3. Выполняя операции, аналогичные поиску в глубину, продвигаемся одновременно по обоим графам от исходных вершин по рёбрам и вершинам с совпадающими метками (метка из графа вопроса должна совпадать с меткой из графа фрагмента).
4. При каждом совпадении ребра/вершины суммируем в общий накопитель баллы совпадения:
  - 4.1. Совпадение рёбер.
    - 4.1.1. Рёбрам разного типа можно присваивать свой вес. Интуитивно, метки AUTHOR, LOK, NAME AGENT должны иметь больший вес, но в окончательных прогонах использовался вес 1 для всех этих типов рёбер.
    - 4.1.2. Некоторые рёбра и вершины разрешается «сокращать»: пропускать при продвижении в глубину в одном графе, не продвигаясь в другом. Например: АСТ, F-АСТ, S-АСТ, МУА.
  - 4.2. Совпадение вершин:
    - 4.2.1. Точное посимвольное совпадение слов – 1 балл
    - 4.2.2. Совпадение лемм – 0.5 балла.
    - 4.2.3. Лемма одной вершины входит в лемму другую как подстрока – 0.5 балла.
5. Накопленная сумма баллов прибавляется к баллу, проставленному предыдущими фильтрами. Заметим, что никакой нормализации баллов здесь не используется, т.к. по построению, шкала схожести с заданным вопросом ограничена размерами графа вопроса и не зависит от фрагмента.

### 3. Эксперименты

В вопросно-ответной дорожке приняли участие две системы. 3 прогона предоставлены в этой работе, ещё один – другим участником. Всего ассессорами были оценены 246 вопросов. Всего для 14 из них был найден хороший ответ (см. Таблица 1). Ещё 60 были отмечены как потенциально имеющие ответ в коллекции.

Вопрос	Фрагмент с ответом
где снимали фильм сталкер?	Солоницын, Николай Гринько, Алиса Фрейндлих, Файме Юрна, Наташа Абр Производство: <b>Россия</b> 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?	<b>Первые дни новорожденные питаются микроводорослями, через 2 3 суток их потихоньку начинают подкармливать разведенными пекарскими дрожжами, живой «пылью»</b>
что такое акдс?	<b>АКДС адсорбированная коклюшно дифтерийно столбнячная вакцина. Адсорбированная коклюшно дифтерийно столбнячная вакцина (АКДС). Прогрессирующие заболевания</b>
"кто входит в совет федерации?	<b>С 1995 г. в Совет Федерации входили губернаторы и спикеры законодательных собраний субъектов РФ. Депутат ГД Алексей Розуван и член Совета Федерации Андрей Ишук подго</b>
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?	<b>23 февраля Ā не красный день календаря.. Не красный, звать, не выходной.. Не красный, нет, он золотой. 23 ФЕВРАЛЯ ОТ А ДО Я! День рождения.</b>
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
биография чингиз айтматов умер где?	<b>Чингиз Торекулович Айтматов. Чингиз Торекулович Айтматов родился в 1928 г. в кишлаке Шекер в Киргизии. Наиболее известные работы Айтматова:</b>
где сейчас играет футболист марат измайлов?	Полузащитники — Динияр Билялетдинов, Марат Измайлов, Дмитрий Лоськов (все — «Локомотив» <b>Москва</b> ), 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 <b>ОВЕН</b> 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?	<b>Экспедитор обязан контролировать силами водителя процесс погрузки (выгрузки), включая пересчет грузовых мест, внешнее состояние упаковки, порядок погрузки, крепл</b>
2. в какой стране изобрели панамы?	116 лет с 1337 по 1453 год.2) Панамы изобрели в <b>Эквадоре</b> .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?	<b>Знаменитая статуя скульптора Мирона изображает атлета, готовящегося к метанию диска. Знаменитая статуя скульптора Поликлета (датированная второй половиной 5 в. до</b>
15 апреля кто по гороскопу?	Телевидение Энциклопедия / Роды / Хотите верить... Информация для авторов ЗДЕСЬ Психотип " <b>Овен</b> " Гороскоп для малышей родившихся с 21 марта до 20 апреляПсихотип "Телец" Гороскоп для малышей родившихся с 21 апреля до 20 маяПсихотип

**Таблица 1 Вопросы с хотя бы одним ответом "good" (всего 14!)**

В таблице во втором столбце представлены расширенные ответы. Полуужирным шрифтом отмечен краткий ответ во фрагменте. Судя по таблице релевантности, у ответов системы xxx-4 краткие ответы совпадают с расширенными ответами, т.е. в таблице весь фрагмент отмечен полуужирным шрифтом. Как видно, в таблице представлены 6 ответов от системы, описанной в этой статье, и 8 ответов от другого участника дорожки. Заметим первый неожиданный результат: пересечений между двумя системами в этом списке нет, хотя система xxx-4 давала ответы на вопросы всех типов, а не только кто/где.

### **3.1 Конфигурации прогонов**

На РОМИП'2010 были сданы 3 прогона системы. Они отличались только наборами каскадных фильтров.

#### **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. Т.е. на Рис. 6 работал только первый блок – A41 «Поиск ответа в тексте вопроса».

#### **Прогон без семантической валидации *naïve***

Во второй конфигурации были включены три фильтра (Рис. 6):

- A41 «Поиск ответа в тексте вопроса»
- A42 «Поиск сочетаний опора-ответ в коллекции»
- A43 «Проверка флуда»

#### **Прогон с семантической валидацией *seman***

Третий прогон был выполнен с полным набором фильтров.

### **3.2 Описание метрик**

Для четвёрки <вопрос, краткий ответ, длинный ответа, документ> ассессоры выставляли четыре оценки:

1. качество вопроса: *good, long, bad\_nonfact, bad*
2. соответствие краткого ответа вопросу: *good, long, suboptimal, partial, miss, bad*
3. соответствие длинного ответа (фрагмента/контекста/сниппета) вопросу: *good, partial, related, bad*
4. соответствие документа вопросу: *good, topic, partial, bad*

В данной работе автор считает важным точность краткого вопроса, поэтому предоставленные организаторами метрики соответствующие длинным ответам здесь не обсуждаются.

Организаторы не предоставили классических метрик полноты/точности/ошибки/аккуратности, поэтому ниже предложены формулы для их расчёта.

Пусть  $QPA$  - QueriesWithPotentialAnswer – множество запросов, у которых есть хотя бы один приблизительный ответ или документ помечен как *good*.  $QPA$  посчитано для общего котла вопросов. Каждый прогон затронул своё подмножество запросов в  $QPA$ , поэтому можно определить множества  $QPA^{lucene}$ ,  $QPA^{naive}$ ,  $QPA^{seman}$ ,  $QPA^{xxx4}$ . Определим несколько вариантов *Полноты* и *Ошибки*:

1. *Полнота* покрытия множества  $QPA$  ответами с оценкой *good*:

$$R_{qpa-good}^i = \frac{|A_{good}^i|}{|QPA|}, \text{ где } A_j^i - \text{множество вопросов, имеющих}$$

в прогоне  $i$  хотя бы один ответ с оценкой  $j$ .

2. *Полнота* покрытия множества  $QPA$  ответами с оценками *good*, *long* или *partial*:

$$R_{qpa-glp}^i = \frac{|A_{good}^i \cup A_{long}^i \cup A_{partial}^i|}{|QPA|}$$

3. *Ошибка*, вычисленная по отношению к общему множеству запросов, учитывая только оценку *good*:

$$E_{total-good}^i = \frac{|R^i \setminus A_{good}^i|}{|U|}, \text{ где } R^i - \text{множество ответов, на}$$

которое был дан хоть один ответ в прогоне  $i$ ,  $U$  – множество всех вопросов, участвующих в дорожке.

4. *Ошибка*, вычисленная по отношению к общему множеству запросов, учитывая оценки *good*, *long* или *partial*:

$$E_{total-glp}^i = \frac{|R^i \setminus A_{good}^i \setminus A_{long}^i \setminus A_{partial}^i|}{|U|}$$

5. *Ошибка*, вычисленная по отношению к множеству *QueriesWithPotentialAnswer*, учитывая только оценку *good*:

$$E_{qpa-good}^i = \frac{|A_{good}^i|}{|QPA^i|}, \text{ где } A_{good}^i \in QPA^i$$

Метрики  $E_{total-good}^i$  и  $E_{total-glp}^i$  поощряют систему за то, что она не выдаёт ответы на вопросы вне множества  $QPA$ , т.е. для тех вопросов, для которых правильных ответов ассессоры так и не нашли. Метрика же  $E_{qpa-good}^i$  игнорирует факты выдачи ошибочных ответов вне множества  $QPA$ .

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	Agpl - 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.

**Таблица 2 Метрики для результатов всех прогонов**

### 3.3 Результаты

Таблица 2 содержит метрики для всех прогонов. Для демонстрации смысла метрик полноты и ошибки в таблицу добавлены искусственные прогоны:

- *noanswer* – прогон гипотетической системы, не давшей ни одного ответа.
- *ideal* – прогон гипотетической системы, давшей правильные ответы на только 14 известных вопросов (см. Таблица 1).

Заметим, что наше гипотетическая идеальная система не даёт 100% полноты. Это связано с тем, что даже после окончания работы ассесоров мы не знаем ответов на 60-14=46 вопросов. Мы только предполагаем, что для этих вопросов можно найти верные ответы в коллекции.

Ожидаемым результатом эксперимента является уменьшение ошибки по мере усложнения конфигурации системы от *lucene->naive->seman*. Улучшение точности при применении семантического фильтра обусловлено значительным уменьшением ложной выдачи ответов, на вопросы, для которых ответа не ожидается вовсе. Прогон *seman* дал ответы на 68 вопросов, против 152 в прогоне *lucene*, срезав почти половину ложных ответов. Однако семантический фильтр также потерял 2 правильных ответа, уменьшив полноту с 8% до 5%.

### Заключение

В результате работы 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 коллекцию из 1.5 млн. документов.

Применение семантического фильтра позволило уменьшить уровень ошибок с 60% до 26%, при ухудшении полноты с 8% до 5%. К сожалению, корректное сравнение с другой системой, участвовавшей в дорожке, затруднено, т.к. эта система не пытается выделить краткие ответы из фрагмента текста.

Дальнейшие исследования видятся в следующих направлениях:

- а) уменьшение ложного срабатывания семантического фильтра;
- б) улучшение реализации задачи NER;
- в) поддержка других типов вопросов, например, определений;
- г) организация повторяемого эксперимента на основе полученной таблицы релевантности по примеру CLEF AVE [3].

## Литература

- [1] Abraham Ittycheriah. A Statistical Approach For Open Domain Question Answering // Advances in Open Domain Question Answering. Springer Netherlands, 2006. Part 1. Vol.32.
- [2] Burger, J. и др. Issues, tasks and program structures to roadmap research in question & answering (Q&A). NIST DUC Vision and Roadmap Documents, 2001. URL: <http://www.nlpir.nist.gov/projects/duc/roadmapping.html>
- [3] Rodrigo, Á., Peñas, A., and Verdejo, F. 2009. Overview of the answer validation exercise 2008. In Proceedings of the 9th Cross-Language Evaluation Forum Conference on Evaluating Systems For Multilingual and Multimodal information Access (Aarhus, Denmark, September 17 - 19, 2008). C. Peters, T. Deselaers, N. Ferro, J. Gonzalo, A. Peñas, G. J. Jones, M. Kurimo, T. Mandl, and V. Petras, Eds. Lecture Notes In Computer Science. Springer-Verlag, Berlin, Heidelberg, 296-313.
- [4] Ellen M. Voorhees. Evaluating Question Answering System Performance // Advances in Open Domain Question Answering. Springer Netherlands, 2006. Part 5. Vol.32.
- [5] The Ephyra Question Answering System [Электронный ресурс]. URL: <http://mu.lti.cs.cmu.edu/trac/Ephyra/wiki>
- [6] Magnini, B., Negri, M., Prevete, R. and Tanev, H., (2002) “Is It the Right Answer? Exploiting Web Redundancy for Answer Validation”. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-2002), Philadelphia, PA
- [7] LASSO: A Tool for Surfing the Answer Net by Dan Moldovan Sanda, Dan Moldovan, A Harabagiu, Marius Poca, Rada Mihalcea, Richard Goodrum, Roxana Girju, Vasile Rus, 1999
- [8] Lucene.NET web site, 2010. <http://lucene.apache.org/lucene.net/>
- [9] Regular Expression Language Elements, MSDN, 2010 <http://msdn.microsoft.com/en-us/library/az24scfc.aspx>
- [10] Автоматическая Обработка Текста [Электронный ресурс]. URL: <http://aot.ru>
- [11] Российский семинар по оценке методов информационного поиска РОМИП, 2010. <http://romip.ru>
- [12] А.А.Соловьёв, О.В.Пескова. Построение вопросно-ответной системы для русского языка: модуль анализа вопросов // Новые информационные технологии в автоматизированных системах: материалы 13 научно-практического семинара.- Моск. гос. ин-т электроники и математики.- 2010.- с.41-49. URL: <http://nps.itas.miem.edu.ru/2010/sbornik13.pdf>



## **Who is to blame and Where the dog is burried? Method of answeres validations based on fuzzy matching of semantic graphs in Question answeting system**

© Alexander Solovyev  
Bauman Moscow State Technical University  
[a-soloviev@mail.ru](mailto:a-soloviev@mail.ru)

Discuss experiments on Question Answering track at Russian Information Retrieval Evaluation Seminar 2010 (ROMIP). The constructed research QA engine searches for answers on questions like “who?” and “where?” in full-text collection. It is based on analysis of semantic relations graphs of in question and text snippet which contains answer. Author follows generic architecture of QA system built-up on generic search engine. Experiments show that applying of proposed semantic filter leads error to decrease from 60% to 26%, while trading for recall decrease from 8% to 5%.