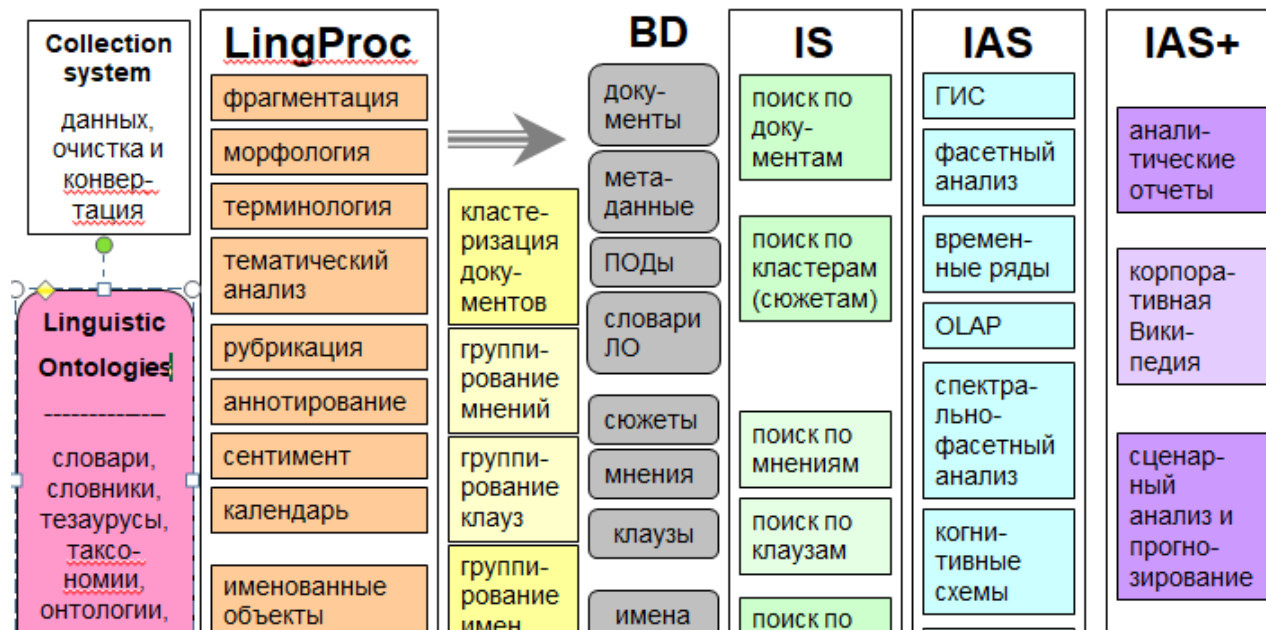


Автоматическое извлечение оценочных отношений на основе нейронных сетей

Наталья Лукашевич
МГУ имени М.В. Ломоносова

Исследовательская группа в МГУ

- Основные интересы:
 - текстовая аналитика и информационно-аналитические системы, представление знаний, анализ тональности
- Имеющиеся ресурсы и технологии:
 - Собственный поисковый движок, онтологии для автоматической обработки текстов, полная линейка автоматической обработки текстов



Распознавание тональности в диалоге

- Empathetic chat-bot=> задача анализа тональности
- Анализ тональности: проблемы
 - Разнообразиие форм выражения позитивных и негативных эмоций
 - Использование отрицаний и других модификаторов тональности
 - Зависимость от предметной области
 - Ирония и сарказм и др.
- В данном докладе
 - Влияние «оценочных» отношений на извлечение тональности

План презентации

- Примеры влияния оценочных отношений на извлечение тональности
- Специальный словарь для извлечения оценочных отношений
- Извлечение отношений по размеченному корпусу на основе нейронных сетей
- Подход на основе опосредованного обучения (distant supervision) для улучшения качества извлечения

Оценочные отношения

- Оценочные отношения – это отношения между двумя сущностями в тексте, которые сообщают нам о позиции сторон по отношению друг к другу
- Оценочные отношения сложным образом взаимодействуют с отношением автора к упоминаемым сущностям
 - Автор может передать свое отношение к одному субъекту через изложение позиций к нему других субъектов
 - Автор может долго перечислять позиции сторон, а затем сделать противоположный вывод

Пример оценочного отношения

- Команда Финляндии обыграла российскую команду
 - Сама новость выглядит просто фактом
 - Слово «обыграть» выглядит позитивным,
 - Но ситуация позитивная только для Финляндии, но не для России, поскольку для российской команды есть негативный эффект
 - Если мы болельщики российской команды (позиция читателя), то для нас сообщение несет отрицательную тональность
- Ср. Сборная Сан-Марино обыграла сборную Лихтенштейна

Оценочные отношения-2

- Председатель одобрил это безобразие
 - Рядом оценочное слово одобрить, однако здесь явное отрицательное мнение автора
- Председатель осудил это безобразие
 - Только отрицательные слова около председателя, но отношение автора позитивное
- Как это можно учесть
 - Словарь? Какой структуры?
 - Разметка обучающей коллекции для машинного обучения? - достаточно сложна, трудно много разметить

Оценочные фреймы

- Создание словаря RuSentiFrames
- Для конкретного слова-предиката описывают
 - позитивные и негативные отношения между участниками
 - Позитивные и негативные эффекты на участников
 - Позитивное/негативное настроение
 - Отношение автора к участникам
- В настоящее время: более 300 фреймов, 7 тысяч слов и выражений

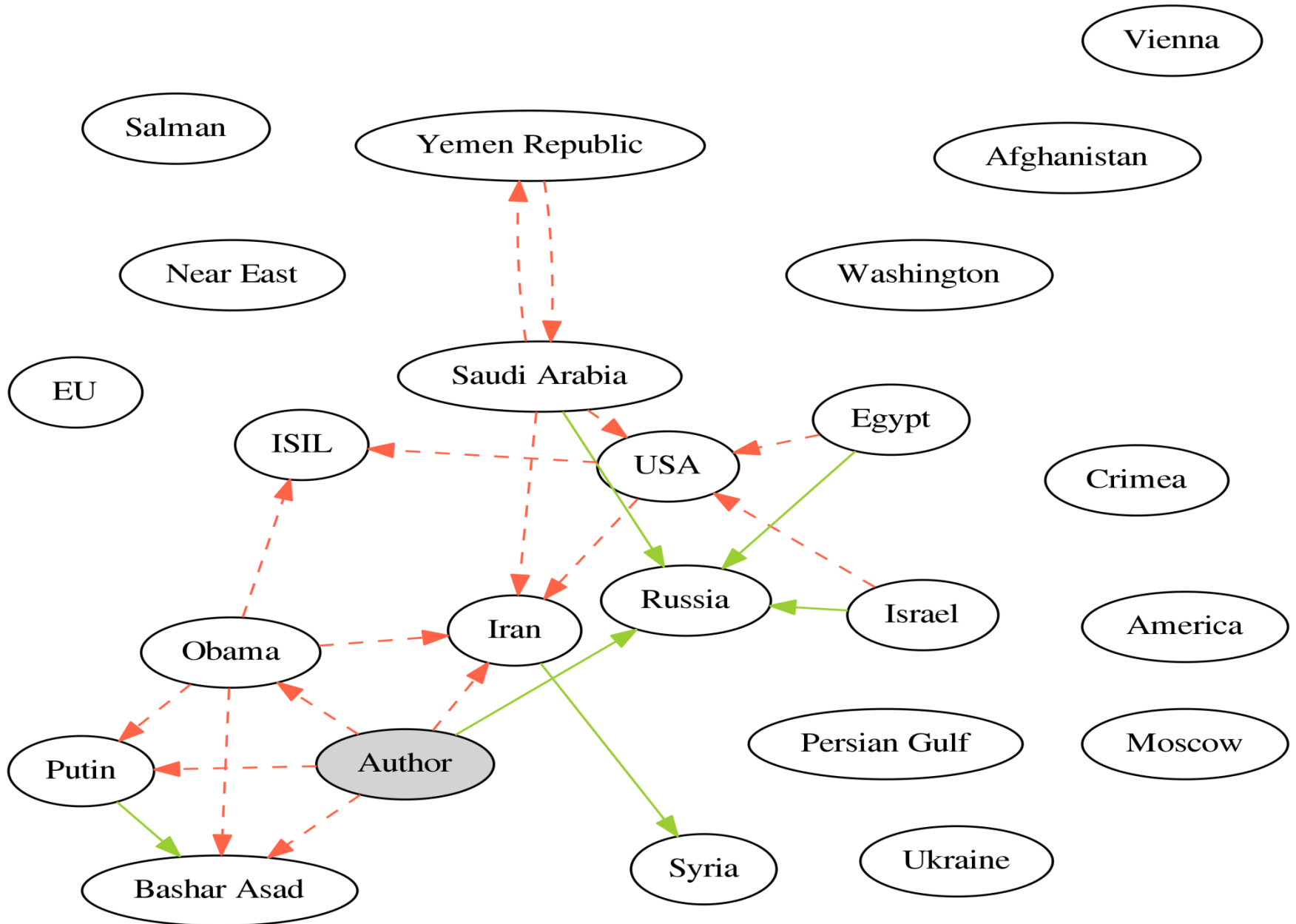
Оценочные фреймы: текущая структура

- Title: завышать
- Comments:
- Variants: завысить, завышать, завышение, переоценивание, переоценивать, переоценить, переоценка, преувеличение, преувеличивать, преувеличить, приукрасить, приукрашать, приукрашивание, приукрашивать
- Roles:
 - «A0»: "агент, тот, кто завышает",
 - «A1»: "объект, то, что завышают«
- Frame:
 - Polarity (A0, A1, pos, 1)
 - Polarity (author, A0, neg, 0.7)
 - Effect (A1, +, 0.7)

Оценочные отношения в новостных и аналитических текстах

- Большой интерес по извлечению мнения автора по отношению к различным субъектам
 - Задача мониторинга
- Но: представлено множество мнений, разные источники и объекты мнений
 - Мнение автора,
 - Оценочные отношения между упомянутыми субъектами
 - Цитируемые мнения третьих лиц
 - Позитивные и негативные события
 - Большое количество сущностей, по отношению к которым нет никаких оценок


Картина целого текста



Разметка коллекции аналитических статей

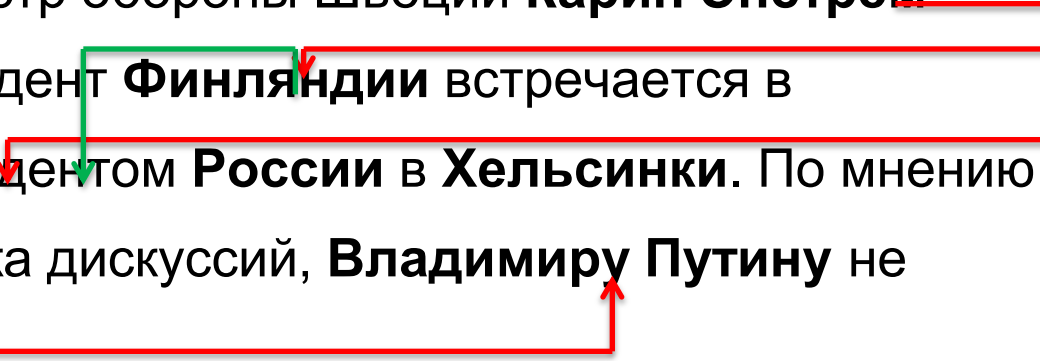
- Русские переводы иностранных аналитических статей с портала inosmi.ru
 - Статьи содержат большое количество мнений
 - Всего 73 статей
 - Корпус RuSentRel
- Подокументная разметка
 - Отношений автора к упоминаемым именованным сущностям
 - Отношений именованных сущностей между собой
 - Более 1500 позитивных и негативных оценок, более 8000 нейтральных

Пример отношений



Отношения **Финляндии** и **Швеции** можно считать хорошими. Ведь входили же страны в состав одного королевства до 1809 года. Страны объединяет также и то, что они не входят в **НАТО**, но являются партнерами альянса. Кроме того, **Финляндия** и **Швеция** укрепляют двустороннее сотрудничество в области обороны.

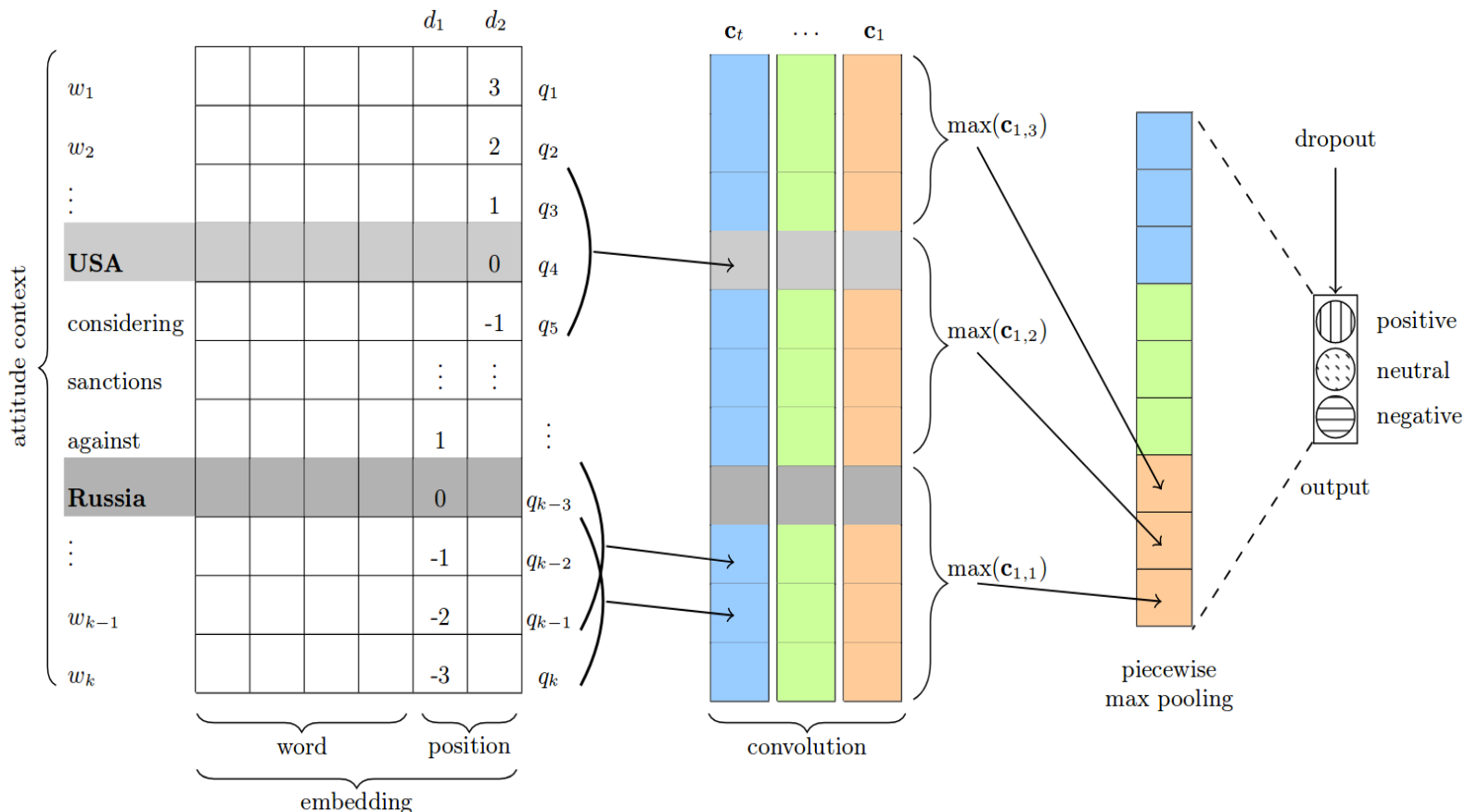
Несмотря на все это, в ходе обсуждений в **Култаранте** возникли разногласия. Бывший министр обороны Швеции **Карин Энстрём** возмутилась тем, что президент **Финляндии** встречается в следующем месяце с президентом **России** в **Хельсинки**. По мнению второго шведского участника дискуссий, **Владимиру Путину** не были бы рады в **Швеции**.



Разметка целого текста

- Пример разметки
 - Обама, Асад, neg
 - США, ИГИЛ, neg
 - Иран, Асад, pos
 - США, Ирак, neg
 - США, Афганистан, neg
 - Япония, США, pos
- Задача
 - Классификация отношений между сущностями на 3 класса
 - Мера качества: F-мера – усреднение между положительным и отрицательным классами
 - Лучший традиционный метод (Random Forest) - 0.27 F-меры
 - Согласие между разметчиками - 0.55

Нейронная сеть Piecewise CNN (PCNN)

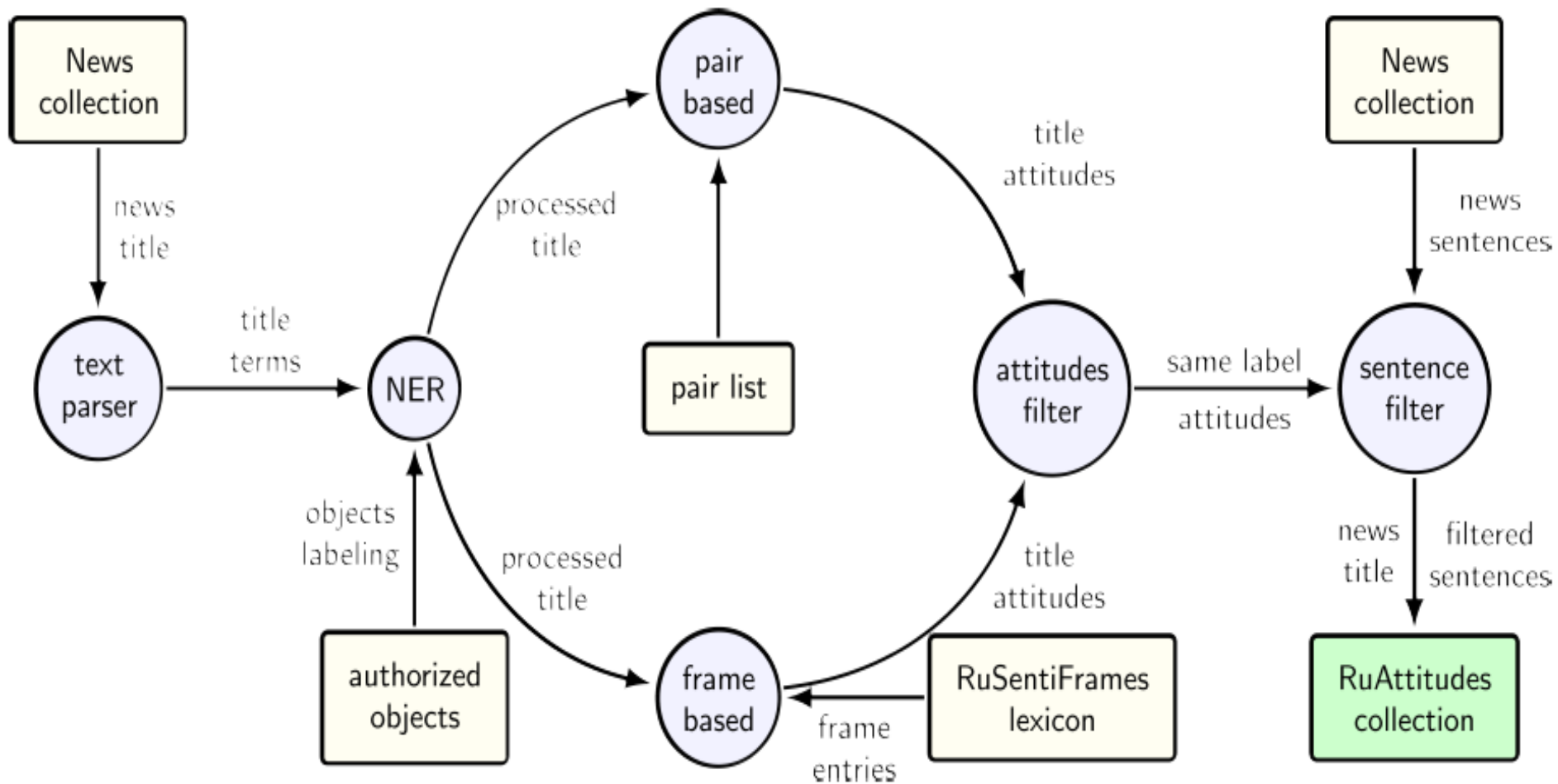


Лучший результат – 0.33 F-меры

Автоматическая разметка текстов (Distant Supervision)

- Идея: обучить на неразмеченных новостных статьях
- Применить к заголовкам статей два критерия, которые должны совпасть
 - Именованные сущности (страны, главы государств) с известным отношением друг к другу (Белоруссия, Россия, pos)
 - Между именованными сущностями должно стоять слово из фреймов с соответствующим прописанным отношением

Схема автоматической разметки отношений



Автоматическая разметка корпуса отношениями (корпус RuAttitudes)

Типы аннотирования	Число извлеченных отношений на уровне текстов	Число текстов, из которых были извлечены отношения	Число извлеч. предложений, включая заголовки
Разметка на основе пар	60788	52377	136496
Разметка на основе фреймов	55566	43383	104205
Пересечение по парам	22476	20885	50958
Пересечение по парам и тональности	14584	13450	33019

Автоматически размеченные примеры негативной и позитивной тональности отношений добавляются к обучающей выборке

Результаты обучения

(на основе извлечения дополнительных отношений)

Модель	F-мера	Точность (P)	Полнота (R)
DS_PCNN	0.41	0.42	0.42
DS_CNN	0.40	0.36	0.46
DS_LSTM	0.36	0.33	0.40
DS_BILSTM	0.35	0.33	0.38
PCNN	0.33	0.48	0.25
CNN	0.33	0.46	0.25
LSTM	0.32	0.48	0.24
BI_LSTM	0.32	0.44	0.25

Заключение

- Тональность в тексте может быть выражена неявным образом: оценочные отношения
- Разметка данных для машинного обучения достаточно сложна
- Создан словарь RuSentiFrames, описывающий предполагаемые отношения и эффекты, связанные с конкретными словами и выражениями
- Словарь был использован для автоматического расширения обучающей коллекции в задаче извлечения оценочных отношений из аналитических статей
- Автоматическая разметка тональности для диалога?