

Искусственный интеллект в цифровых гуманитарных исследованиях

Воронцов Константин Вячеславович

д.ф.-м.н., проф. РАН, заведующий лабораторией Машинного Интеллекта МФТИ

Карабулатова Ирина Советовна

д.филол.н., проф., профессор-исследователь РУДН

Примеры социально значимых проектов на основе технологий ИИ и понимания естественного языка

1. Конкурс ПРО//ЧТЕНИЕ

автоматический поиск смысловых ошибок в школьных сочинениях

2. «Мастерская знаний»

составление и анализ тематических подборок научных статей

3. Анти-фейк и анти-постправда

выявление потенциально опасного дискурса в СМИ и соцмедиа

4. Управление молодёжной политикой

поиск и поддержка позитивных инициатив и активностей

Примеры социально значимых проектов на основе технологий ИИ и понимания естественного языка

1. Конкурс ПРО//ЧТЕНИЕ

автоматический поиск смысловых ошибок в школьных сочинениях

2. «Мастерская знаний»

составление и анализ тематических подборок научных статей

3. Анти-фейк и анти-постправда

выявление потенциально опасного дискурса в СМИ и соцмедиа

4. Управление молодёжной политикой

поиск и поддержка позитивных инициатив и активностей

КОНКУРС ПРО//ЧТЕНИЕ

<http://ai.upgreat.one>

ПРО//ЧТЕНИЕ

5

ЗАДАЧА:

Автоматическое выявление смысловых ошибок в текстах на естественных языках: русском и английском.

ТЕХНОЛОГИЧЕСКИЙ БАРЬЕР:

ИИ находит и аннотирует ошибки на уровне специалиста в условиях ограниченного времени.

Для конкурса собирается уникальный датасет сочинений школьников, смысловые ошибки выявляются преподавателями и экспертами ЕГЭ в соответствии со стандартами ФИПИ

ФАКТИЧЕСКАЯ ОШИБКА

автор высказывания А.Франц

В своем высказывании «Если человек зависит от природы, то и она от него зависит» Д. Мережковский говорит о необходимости защиты природы.

ЛОГИЧЕСКАЯ ОШИБКА
тезис не обоснован



ПРИЗОВОЙ
ФОНД

РУССКИЙ
100 МЛН. РУБ.

АНГЛИЙСКИЙ
100 МЛН. РУБ.

Пример. «Проблема ответственности за свои поступки (по тексту В. Ф. Тендрякова)»

Алгоритмическая разметка

Нередко люди совершают плохие поступки, забывая о том, что, даже скрыв свой поступок от других, человек не скроется от своей совести. Что же такое безнравственный поступок? Безнравственный поступок - это поступок, не соответствующий моральным нормам.

Можно ли оправдать безнравственный поступок? Именно эту проблему В. Ф. Тендряков поднимает в своем тексте. Докажем сказанное примерами из представленного отрывка.

В тексте В. Ф. Тендряков говорит о том, что человек во благо себе может легко совершить низкий поступок, не испытав при этом чувство стыда. Человек сможет оправдать свой поступок перед самим собой, объяснив причину. В пример автор приводит поведение героя, который часто в жизни совершал безнравственные поступки. Он врал, дрался и крал. Мы видим, что до войны герой привык совершать плохие поступки. Он всегда оправдывался, потому что не хотел нести ответственность за свои действия, а значит не испытывал мучения совести. Мы знаем, что муки совести - это первое и самое сильное наказание, которое получает человек, совершивший плохой поступок. Но наш герой не получал никакого наказания и поэтому продолжал совершать безнравственные поступки. Проанализировав поведение главного героя, я убедилась в том, что человек обязан нести ответственность за свои поступки всегда, и поэтому я утверждаю, что нельзя оправдывать даже мелкие безнравственные поступки.

СВЯЗЬ Р.ПОВТОР
Р.ПОВТОР Р.ЛИШН ПРОБЛЕМА
Р.ПОВТОР Р.ПОВТОР Р.ПОВТОР
Р.ЛИШН
Р.ПОВТОР
Р.ПОВТОР
Р.ПОВТОР
Р.ПОВТОР Г.ОДНОР Г.ОДНОР Г.ОДНОР
Г.ВИДОВР Р.ПОВТОР
Р.ПОВТОР Р.ПОВТОР
Р.ПОВТОР Р.ПОВТОР
Р.ПОВТОР Г.ВИДОВР Р.ПОВТОР
Р.ПОВТОР
Р.ПОВТОР

Экспертная разметка 2

Нередко люди совершают плохие поступки, забывая о том, что, даже скрыв свой поступок от других, человек не скроется от своей совести. Что же такое безнравственный поступок? Безнравственный поступок - это поступок, не соответствующий моральным нормам.

Можно ли оправдать безнравственный поступок? Именно эту проблему В. Ф. Тендряков поднимает в своем тексте. Докажем сказанное примерами из представленного отрывка.

В тексте В. Ф. Тендряков говорит о том, что человек во благо себе может легко совершить низкий поступок, не испытав при этом чувство стыда. Человек сможет оправдать свой поступок перед самим собой, объяснив причину. В пример автор приводит поведение героя, который часто в жизни совершал безнравственные поступки. Он врал, дрался и крал. Мы видим, что до войны герой привык совершать плохие поступки. Он всегда оправдывался, потому что не хотел нести ответственность за свои действия, а значит не испытывал мучения совести. Мы знаем, что муки совести - это первое и самое сильное наказание, которое получает человек, совершивший плохой поступок. Но наш герой не получал никакого наказания и поэтому продолжал совершать безнравственные поступки. Проанализировав поведение главного героя, я убедилась в том, что человек обязан нести ответственность за свои поступки всегда, и поэтому я утверждаю, что нельзя оправдывать даже мелкие безнравственные поступки.

Р.ПОВТОР Т1
Р.ПОВТОР Т1
Р.ПОВТОР Т2 Р.ПОВТОР
ПРОБЛЕМА Р.ПОВТОР Т
ПРИМЕР Р.ПОВТОР Т3
Р.ТАВТ Т4 Р.ПОВТОР Т1
Р.ПОВТОР Т1
Р.ТАВТ Т4
Р.ПОВТОР Т1
Р.ТАВТ Т4 Р.ПОВТОР Т1
Р.ТАВТ Т4 Р.ПОВТОР Т1
Р.ТАВТ Т4 Р.ПОВТОР Т1
Р.ТАВТ Т4 Р.ПОВТОР Т1
ПОЯСНЕНИЕ
Р.ПОВТОР Т1
Р.ПОВТОР Т1

Примеры социально значимых проектов на основе технологий ИИ и понимания естественного языка

1. Конкурс ПРО//ЧТЕНИЕ

автоматический поиск смысловых ошибок в школьных сочинениях

2. «Мастерская знаний»

составление и анализ тематических подборок научных статей

3. Анти-фейк и анти-постправда

выявление потенциально опасного дискурса в СМИ и соцмедиа

4. Управление молодёжной политикой

поиск и поддержка позитивных инициатив и активностей

Концепция «Мастерской знаний»

«Огромное и все возрастающее богатство знаний разбросано сегодня по всему миру. Этих знаний, вероятно, было бы достаточно для решения всего громадного количества трудностей наших дней, но они рассеяны и неорганизованы. Нам необходима очистка мышления в **своеобразной мастерской**, где можно получать, сортировать, суммировать, усваивать, разъяснять и сравнивать знания и идеи.» – *Герберт Уэллс, 1940*

(An immense and ever-increasing wealth of knowledge is scattered about the world today; knowledge that would probably suffice to solve all the mighty difficulties of our age, but it is dispersed and unorganized. We need a sort of mental clearing house for the mind: a **depot where knowledge and ideas are received, sorted, summarized, digested, clarified and compared** – *Herbert Wells, 1940*)



Сегодня технологии IR/ML/NLP позволяют решать такие задачи

Функции «Мастерской знаний»

Подборка – долгосрочный поисковый интерес пользователя или группы

Поисково-рекомендательные функции:

- поиск тематически близких документов по **подборке**
- мониторинг новых документов по тематике **подборки**

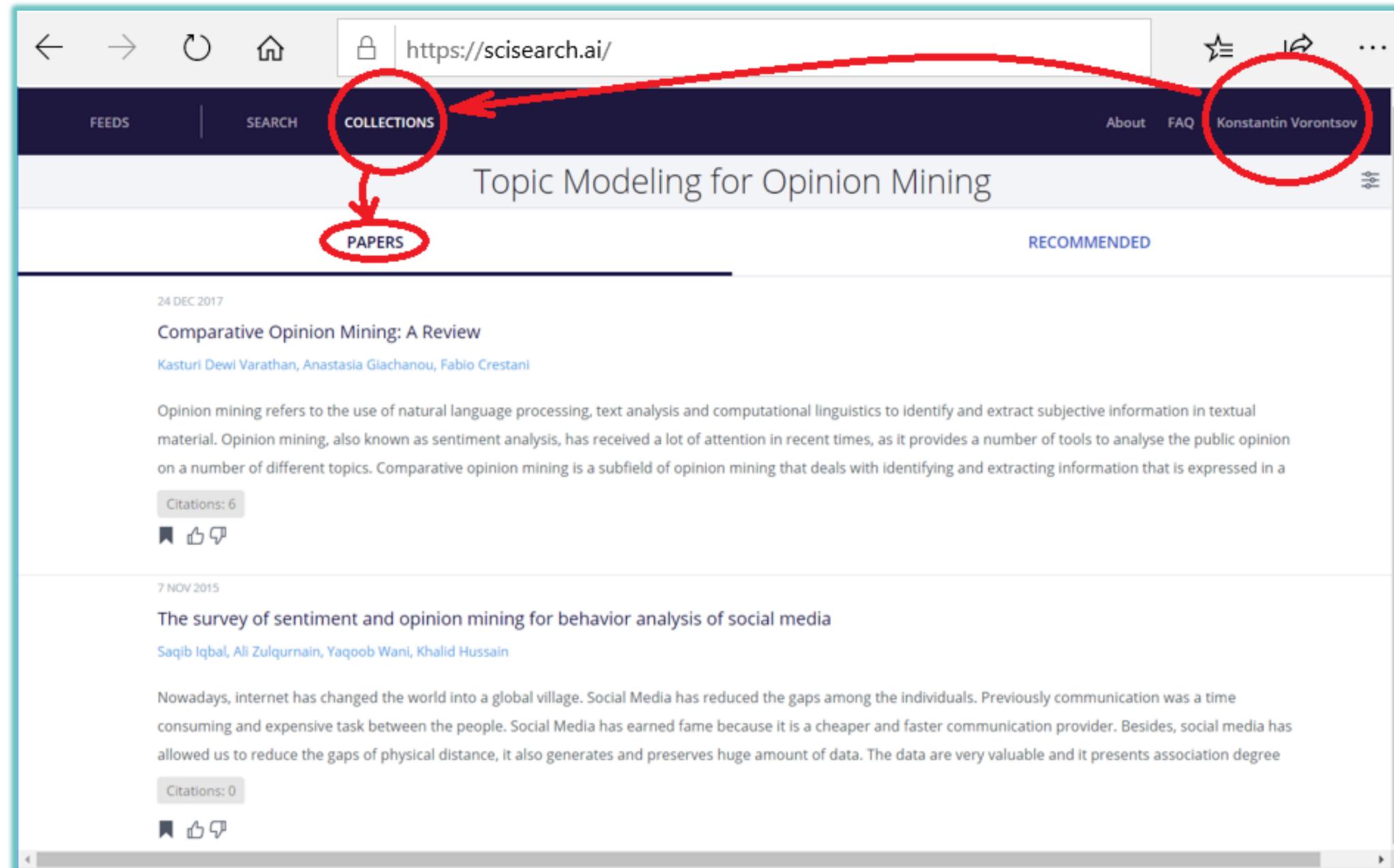
Аналитические функции:

- рекомендация порядка чтения документов внутри **подборки**
- автоматизация реферирования **подборки**
- систематизация тем, идей, решений, мнений внутри **подборки**

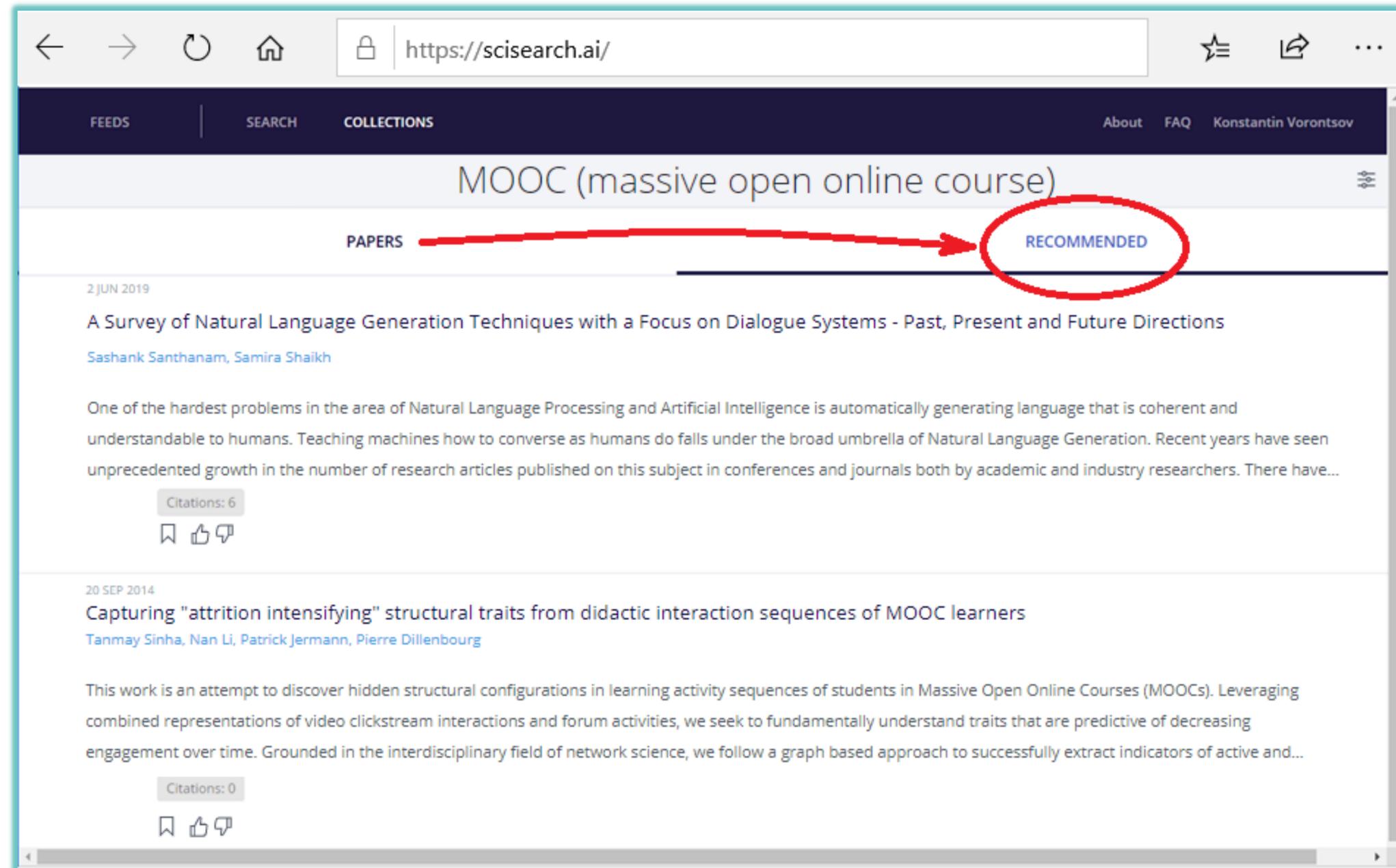
Коммуникативные функции:

- совместное составление, обсуждение, использование **подборок**
- интерактивная визуализация и инфографика по **подборке**

Поиск и рекомендации в SciSearch.ai



Поиск и рекомендации в SciSearch.ai



The screenshot shows the SciSearch.ai website interface. The browser address bar displays <https://scisearch.ai/>. The navigation menu includes FEEDS, SEARCH, and COLLECTIONS. The search results are for the query "MOOC (massive open online course)". Two tabs are visible: "PAPERS" and "RECOMMENDED". The "RECOMMENDED" tab is circled in red, and a red arrow points from the "PAPERS" tab to it. Below the tabs, two search results are displayed:

2 JUN 2019
A Survey of Natural Language Generation Techniques with a Focus on Dialogue Systems - Past, Present and Future Directions
Sashank Santhanam, Samira Shaikh

One of the hardest problems in the area of Natural Language Processing and Artificial Intelligence is automatically generating language that is coherent and understandable to humans. Teaching machines how to converse as humans do falls under the broad umbrella of Natural Language Generation. Recent years have seen unprecedented growth in the number of research articles published on this subject in conferences and journals both by academic and industry researchers. There have...

Citations: 6

20 SEP 2014
Capturing "attrition intensifying" structural traits from didactic interaction sequences of MOOC learners
Tanmay Sinha, Nan Li, Patrick Jermann, Pierre Dillenbourg

This work is an attempt to discover hidden structural configurations in learning activity sequences of students in Massive Open Online Courses (MOOCs). Leveraging combined representations of video clickstream interactions and forum activities, we seek to fundamentally understand traits that are predictive of decreasing engagement over time. Grounded in the interdisciplinary field of network science, we follow a graph based approach to successfully extract indicators of active and...

Citations: 0

Поиск и рекомендации в SciSearch.ai

The screenshot displays the SciSearch.ai website interface. The browser's address bar shows the URL <https://scisearch.ai/>. The navigation menu includes FEEDS, SEARCH, and COLLECTIONS. The main content area features a search result for "MOOC (massive open online course)" by Sashank Santhanam and Samira Shaikh, dated 2 JUN 2019. The article title is "A Survey of Natural Language Generation T...". The abstract begins with "One of the hardest problems in the area of Natural Language Generation is to generate text that is understandable to humans. Teaching machines how to generate text has seen unprecedented growth in the number of research articles published in the field of Natural Language Generation. Recent years have seen both by academic and industry researchers. There have...". The article has 6 citations. A red circle highlights the citation count, and a red arrow points from it to the "Add to collections" dialog box. The dialog box is titled "Add to collections" and contains a list of collection categories: Exploratory Search, MOOC (massive open online course), Opinion Mining and Sentiment Analysis with Topic Modeling, Textual Complexity and Readability, and Topic modeling of genomic data. The "MOOC (massive open online course)" option is selected. A red circle highlights this option, and a red arrow points from it to the "SAVE CHANGES" button. Below the list is a "NEW COLLECTION" link. The background shows a "RECOMMENDED" section with a red circle around the word "RECOMMENDED".

Технология тематического поиска VigARTM

Схема эксперимента:

- длинные запросы (1 стр. А4)
- 100 запросов на коллекцию
- 3 ассессора на каждый запрос
- от 10 до 60 минут на запрос
- разметка на Яндекс.Толока
- две коллекции техно-новостей:



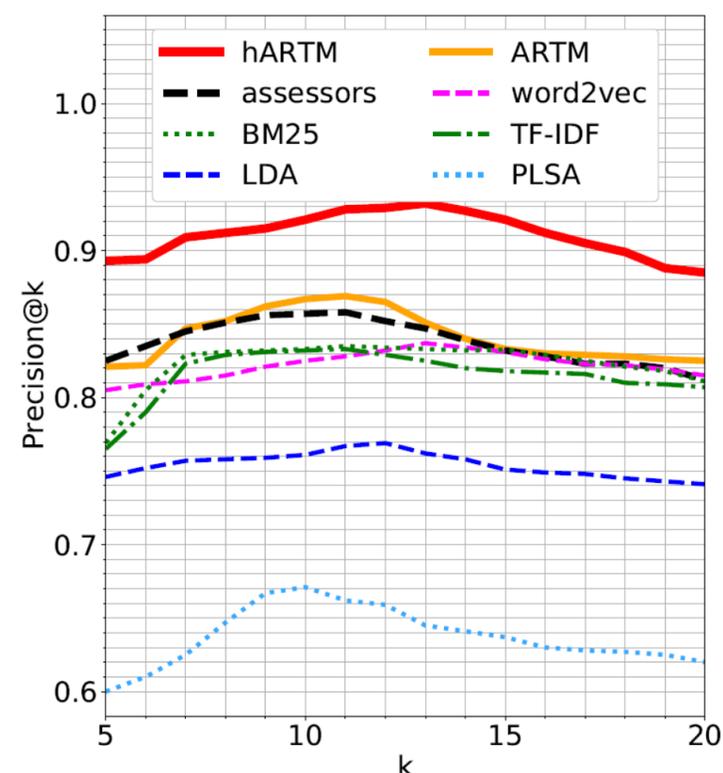
(170K Russian docs)



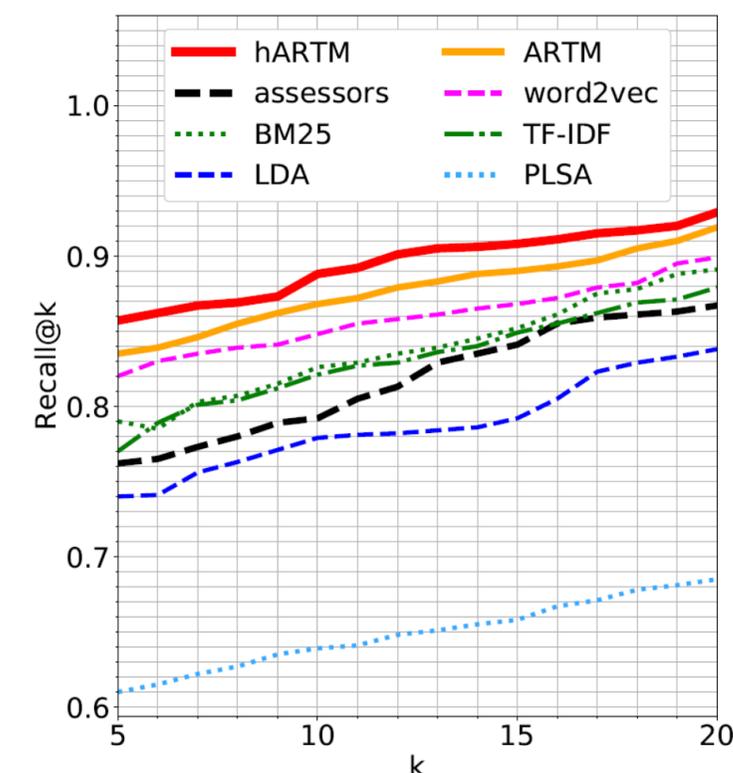
(750K English docs)

Оценки качества поиска:

точность (precision@k)



полнота (recall@k)



Ianina A., Golitsyn L., Vorontsov K. [Multi-objective topic modeling for exploratory search in tech news](#). AINL 2017.

Ianina A., Vorontsov K. [Regularized multimodal hierarchical topic model for document-by-document exploratory search](#). 2019.

Автоматизация реферирования

Основные задачи машинного обучения:

- Формирование обучающей выборки: **paper** → **(refs, survey)**
- Ранжирование статей для сценария реферата
- Выбор релевантных фраз из текста статьи для каждого суфлёра
- Ранжирование выбранных фраз для каждого суфлёра
- Выбор релевантного контекста по данной ссылке, например:

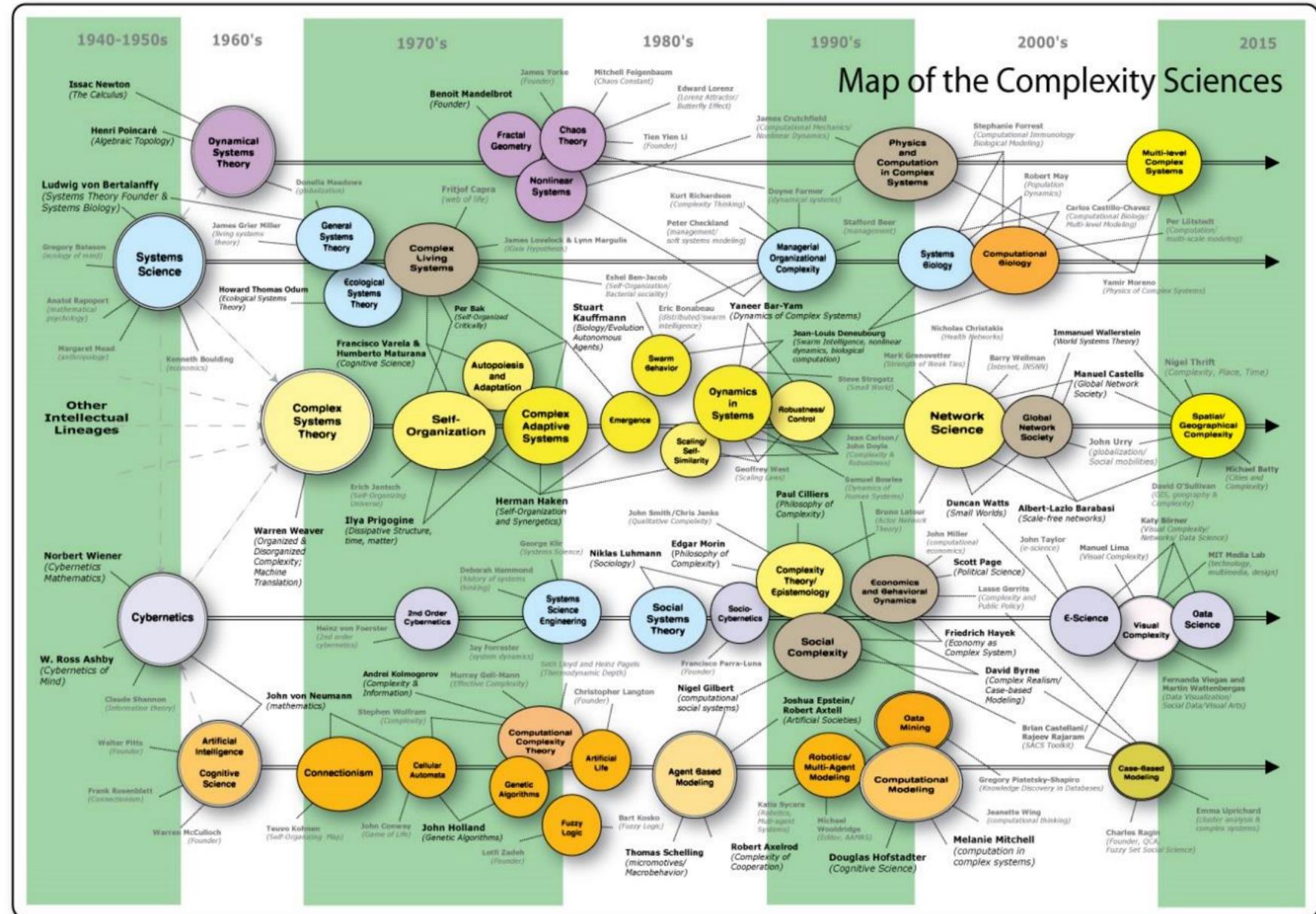
Few contextual citation graphs are publicly available. The [ACL Anthology Network \(AAN\)](#) ([Radev et al., 2009](#)) is one such contextual citation graph built from the [ACL Anthology corpus](#) ([Bird et al., 2008](#)), consisting of 24.6K papers manually augmented with citation information. [CiteSeer](#) ([Giles et al., 1998](#)) provides a large corpus consisting of 1.0M papers with full text and bibliography entries parsed from PDFs. [Saier and Farber \(2019\)](#) introduces a contextual citation graph of approximately 1.0M [arXiv](#) papers with full text LaTeX parses where citations are linked to papers in the Microsoft Academic Graph.

M.Yasunaga, J.Kasai, R.Zhang, A.Fabbri, I.Li, D.Friedman, D.Radev. ScisummNet: A Large Annotated Corpus and Content-Impact Models for Scientific Paper Summarization with Citation Networks. 2019.

Визуализация и дистантное чтение (distant reading)

Осями на карте могут быть:

- время
- спектр тем
- сложность
- обзорность
- актуальность
- «хайповость»
- цитируемость



Примеры социально значимых проектов на основе технологий ИИ и понимания естественного языка

1. Конкурс ПРО//ЧТЕНИЕ

автоматический поиск смысловых ошибок в школьных сочинениях

2. «Мастерская знаний»

составление и анализ тематических подборок научных статей

3. Анти-фейк и анти-постправда

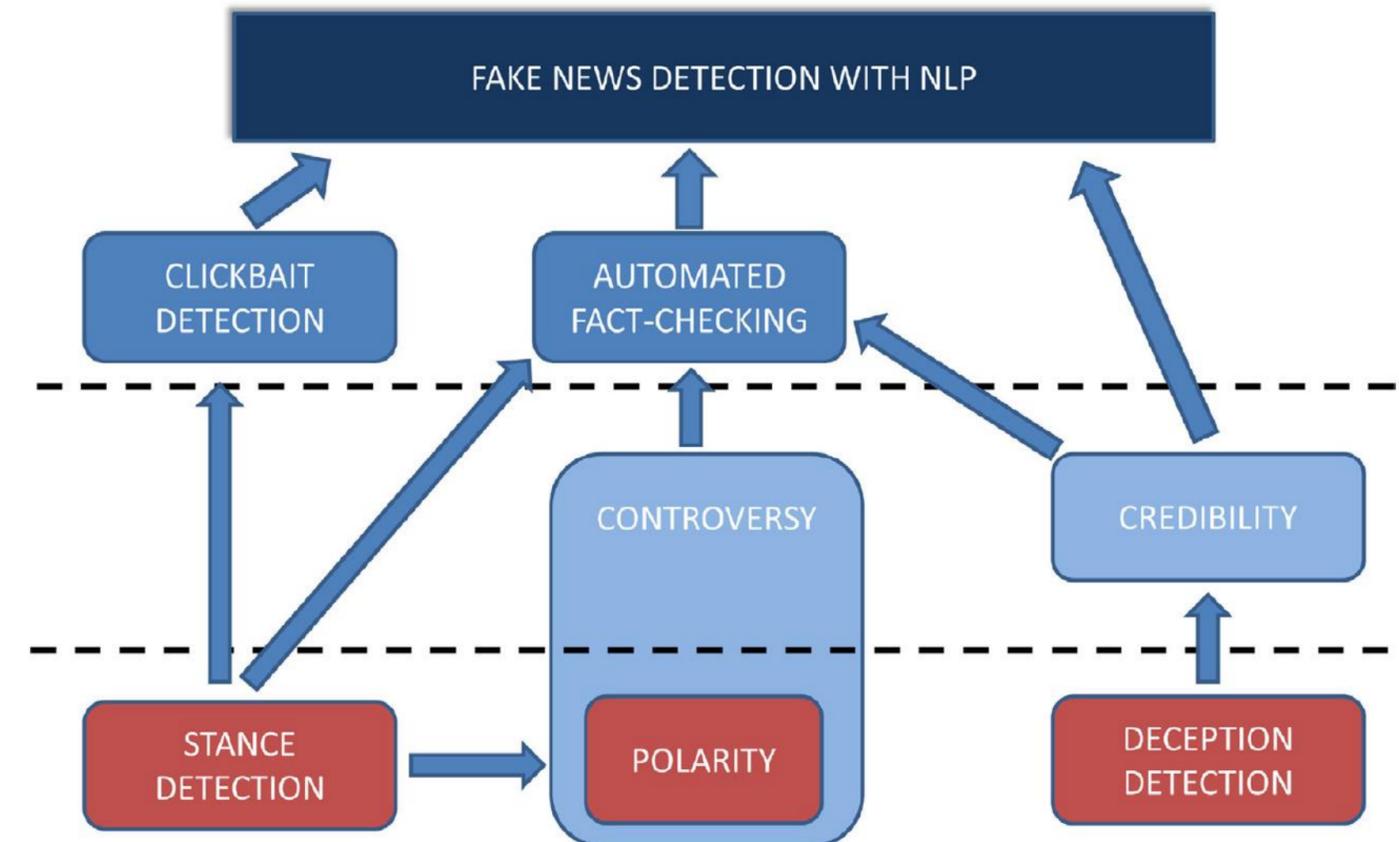
выявление потенциально опасного дискурса в СМИ и соцмедиа

4. Управление молодёжной политикой

поиск и поддержка позитивных инициатив и активностей

Область исследований «Fake News Detection»

1. Deception Detection
выявление обмана в тексте новости
2. Automated Fact-Checking
автоматическая проверка фактов
3. Stance Detection
выявление позиции за/против запроса (claim)
4. Controversy Detection
выявление и кластеризация разногласий
5. Polarization Detection
классификация позиций по многим темам
6. Clickbait Detection
выявление противоречий заголовка и текста
7. Credibility Scores
оценка достоверности источника или новости

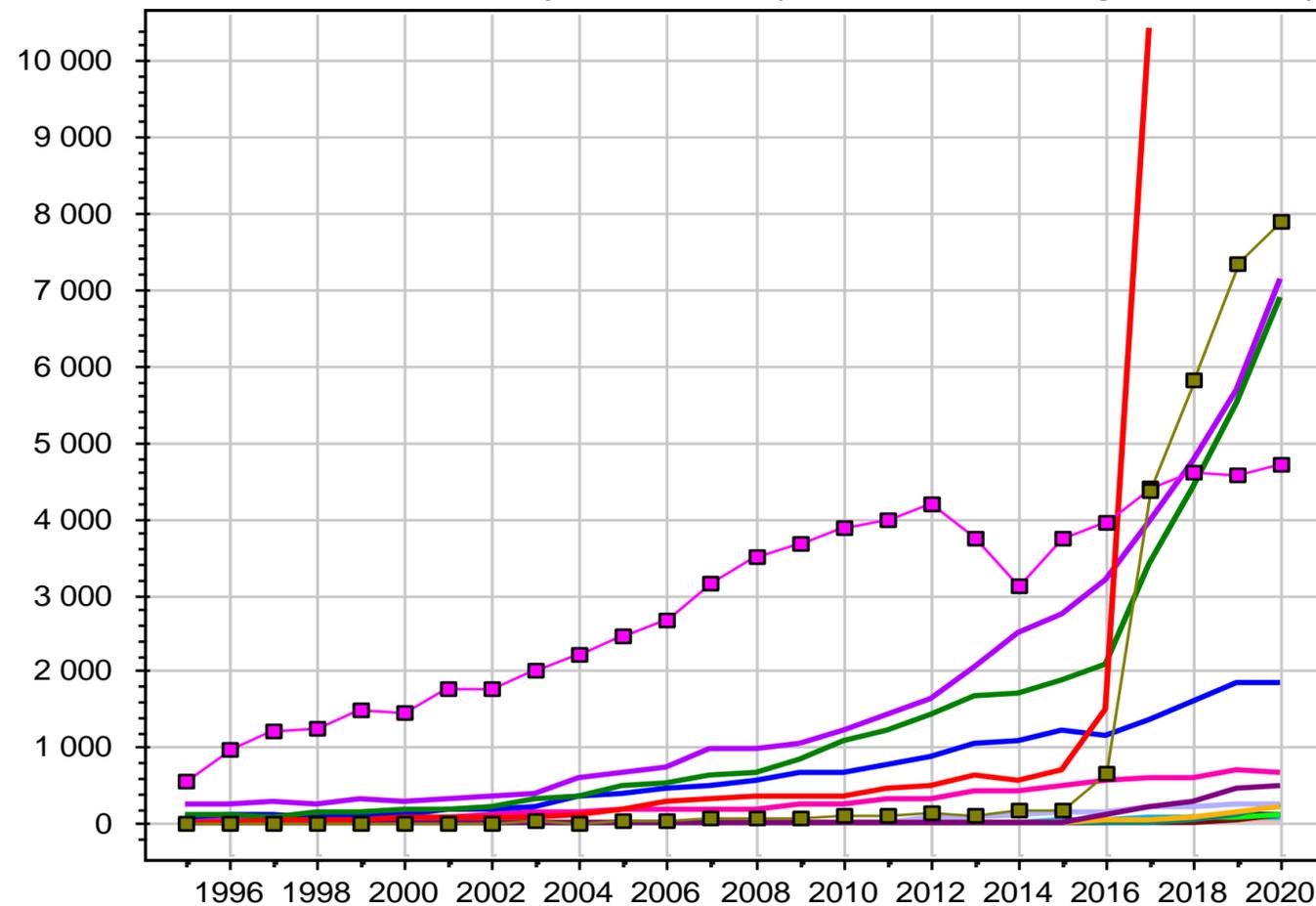


E.Saquete, D.Tomás, P.Moreda, P.Martínez-Barco, M.Palomar. Fighting post-truth using natural language processing: A review and open challenges. Expert Systems With Applications, Elsevier, 2020.

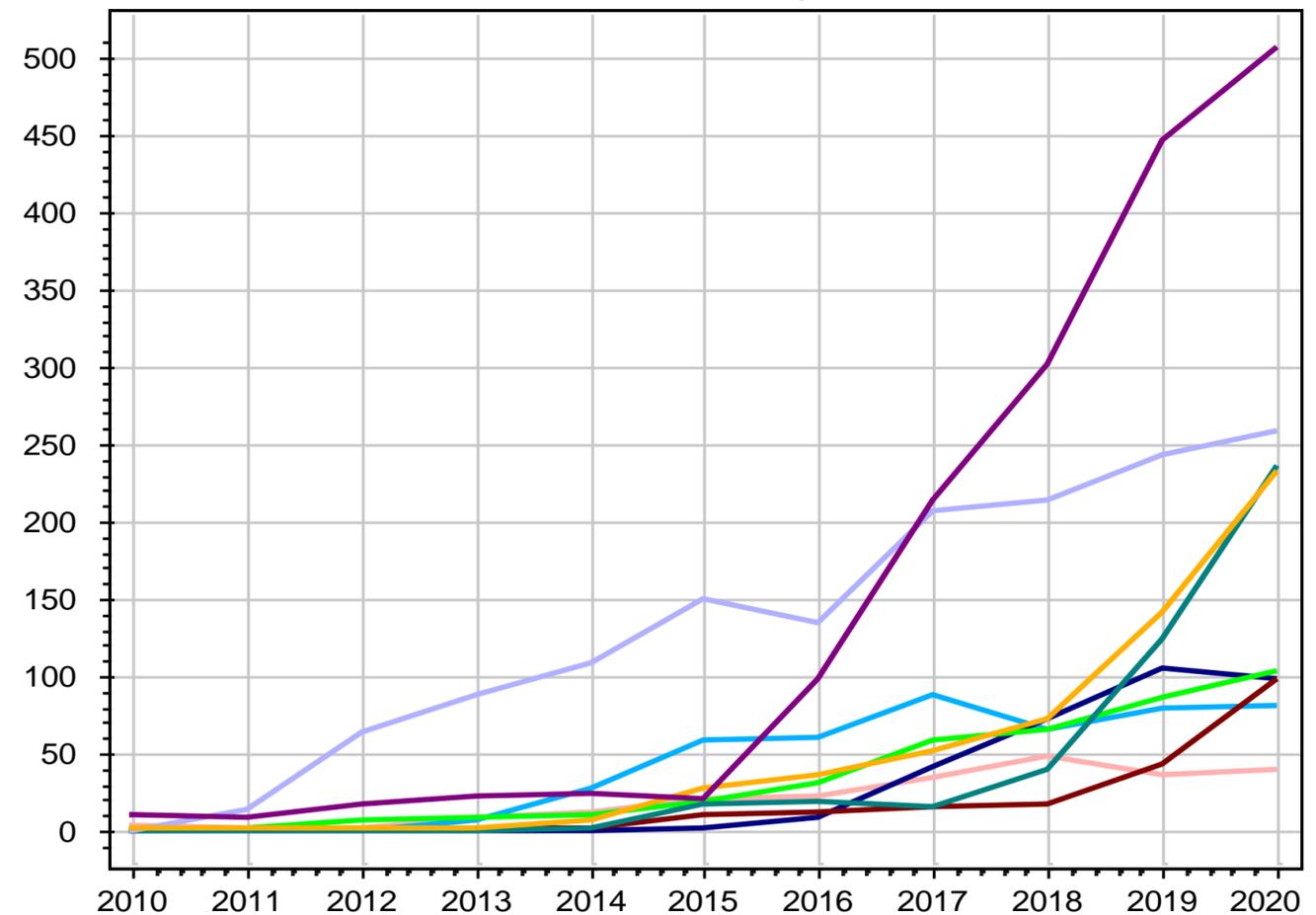
Fake News и близкие темы исследований

Библиометрический анализ по данным Google Scholar:

Число публикаций (по данным Google Scholar)



Новые тренды последних 10 лет



- post-truth
- information warfare
- fake news
- political polarization
- fact checking
- language manipulation
- deception detection
- stance detection
- rumor detection
- misinformation detection
- hoax detection
- propaganda detection
- clickbait detection
- controversy detection
- deceptive opinion spam
- virality prediction

Типология потенциально опасного дискурса и система подзадач ML/NLP для его детекции

воздействия → фейки → пропаганда → инф.война

1. детекция приёмов манипулирования
2. детекция замалчивания
3. детекция обмана (deception detection), слухов (rumors d.), мистификаций (hoaxes d.)
4. детекция кликбэйта (clickbait detection)
5. автоматическая проверка фактов (auto fact-checking)
6. детекция позиции (stance d.), противоречий (controversy d.), поляризации (polarization d.)
7. выявление конструкторов картины мира: идеологем, мифологем
8. оценивание возможных психо-эмоциональных реакций
9. выявление целевых аудиторий воздействия
10. оценивание виральности (virality prediction)
11. оценивание достоверности источников (credibility scores)
12. детекция прямой агрессии (угрозы, призывы, провокации, вербовка, экстремизм)

Четыре основных типа подзадач ML/NLP

- 1. Классификация текста (новости или предложения) целиком**
 - *deception detection, fact-checking, text credibility*
- 2. Классификация пары текстов**
 - *stance, controversy, polarization, clickbait detection*
 - выявление противоречий, разногласий, замалчивания
- 3. Выделение и классификация (тегирование) фрагментов текста**
 - *поиск лингвистических маркеров (linguistic-based cues) в тексте*
 - детекция приёмов манипулирования
 - выявление конструкторов картины мира: идеологем, мифологем
 - выявление психо-эмоциональных реакций и целевых аудиторий
- 4. Кластеризация или тематическое моделирование**
 - *кластеризация мнений по заданной теме (controversy detection)*
 - *выявление устойчивых сочетаний мнений (polarization detection)*
 - выявление мнений как сочетаний слов, их семантических ролей и тональностей
 - выявление «картин мира» – устойчивых сочетаний мнений и идеологем

Задача выделения мнений в теме или событии

... Президент Петр Порошенко заявил, что Россия де-факто конфисковала украинские предприятия, которые находятся на неподконтрольной Киеву территории. Сегодня ДНР и ЛНР "национализировали" украинские предприятия ... При этом Кремль защитил конфискацию предприятий в ЛДНР ... Украина потребует расширить санкции ... За все эти действия обязательно наступит наказание. Украина потребует расширения санкций на тех, кто украл украинские предприятия ... *(Kiev opinion)*

... По словам Захарченко, Киев встретит свой "ужасный конец"... Киев возьмется за ум, и в целях спасения собственной промышленности снимет блокаду ... Обстановка, которую искусственно создала Украина с блокадой Донбасса, вынудила ... кошмарит свой народ ... если в Киеве были приняты какое-либо постановление ... положительные результаты, как в республиках, так и в России... Если им удастся сместить Порошенко и при этом не развалить Украину, то все вернется на свои места ... *(Moscow opinion)*

Subject

Object

Agent

Locative

Negative lexicon

Dependent word

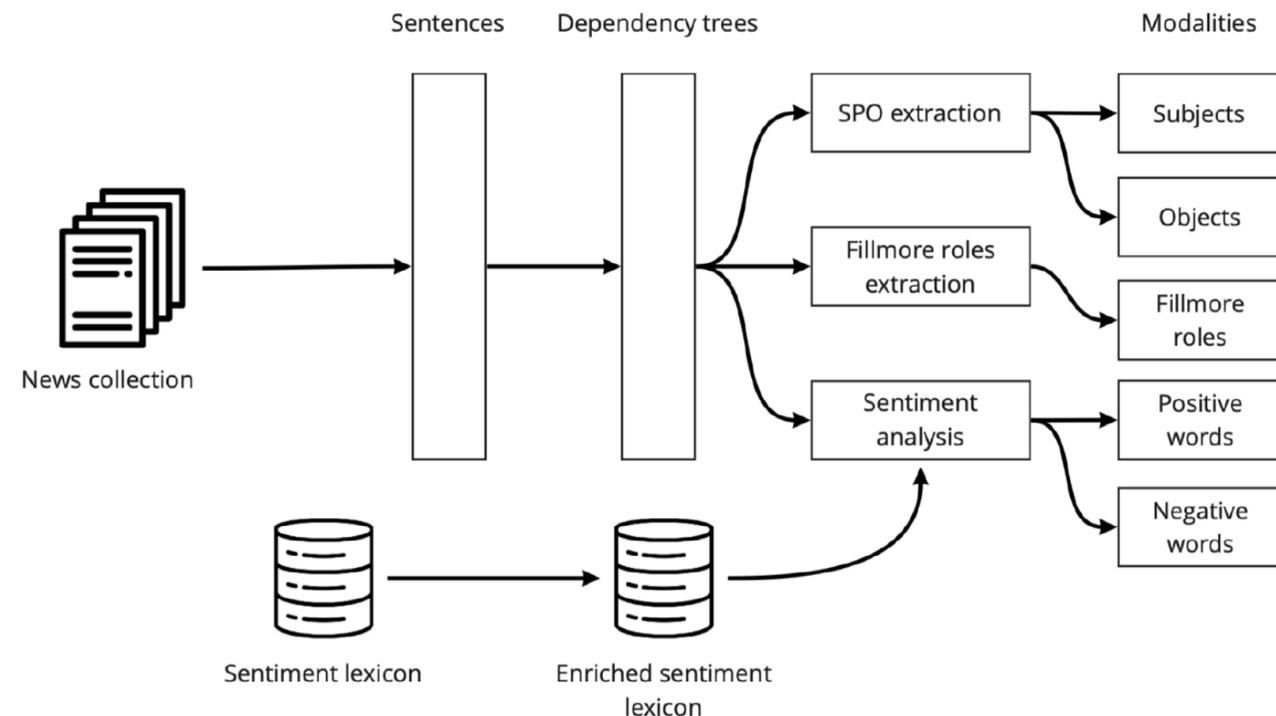
Слова «Порошенко», «Россия», «Украина» встречаются одинаково часто

«Порошенко» — субъект в первом тексте и объект во втором

«Россия» — агент в первом тексте и локация во втором

Негативная тональность: «Россия», «Кремль» в 1-ом, «Киев», «Украина» во 2-ом

Задача выделения мнений в теме или событии



Modalities	<i>Pr</i>	<i>Rec</i>	<i>F1</i>
TF-IDF	0.51	0.95	0.67
SPO	0.59	0.7	0.64
FR	0.86	0.49	0.65
Sent	0.69	0.57	0.66
SPO+FR	0.86	0.68	0.76
SPO+Sent	0.83	0.78	0.81
FR+Sent	0.9	0.52	0.67
All	0.77	0.97	0.86

LPR Business

Modalities	<i>Pr</i>	<i>Rec</i>	<i>F1</i>
TF-IDF	0.57	0.97	0.72
SPO	0.56	0.99	0.72
FR	0.67	0.97	0.79
Sent	0.56	0.55	0.55
SPO+FR	0.72	0.99	0.83
SPO+Sent	0.57	0.99	0.72
FR+Sent	0.73	0.97	0.83
All	0.77	0.94	0.85

Paris Trump

Мнение формализуется как устойчивое сочетание слов, терминов, именованных сущностей, их семантических ролей по Филлмору и их тональных окрасок. Все они используются в тематической модели как отдельные модальности.

Feldman D. G., Sadekova T. R., Vorontsov K. V. [Combining Facts, Semantic Roles and Sentiment Lexicon in A Generative Model for Opinion Mining](#). Computational Linguistics and Intellectual Technologies. Dialogue 2020.

Примеры социально значимых проектов на основе технологий ИИ и понимания естественного языка

1. Конкурс ПРО//ЧТЕНИЕ

автоматический поиск смысловых ошибок в школьных сочинениях

2. «Мастерская знаний»

составление и анализ тематических подборок научных статей

3. Анти-фейк и анти-постправда

выявление потенциально опасного дискурса в СМИ и соцмедиа

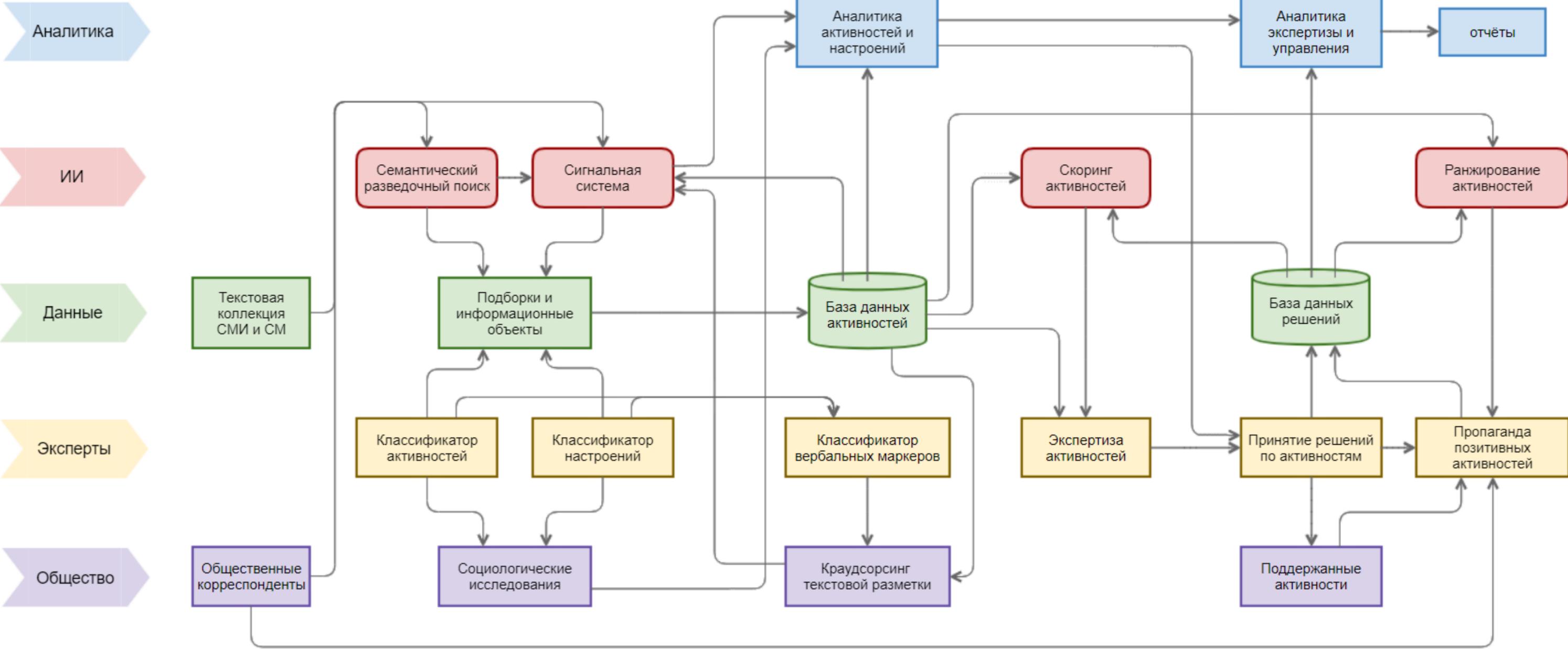
4. Управление молодёжной политикой

поиск и поддержка позитивных инициатив и активностей

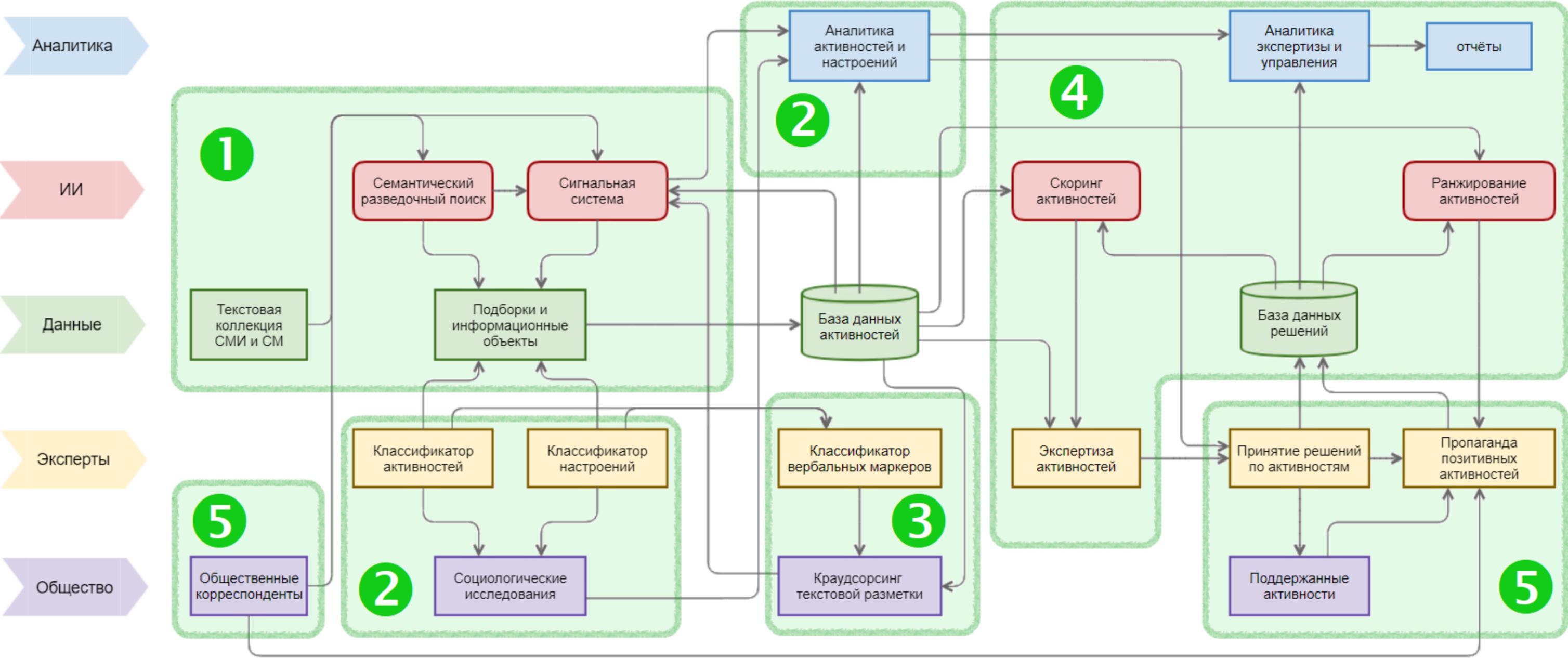
Предпосылки создания системы управления социально-политической активностью молодежи

1. Постоянно возникают новые виды инициатив и активностей, как позитивных, так и негативных
2. Эти активности хорошо заметны в СМИ и соцмедиа
3. Выявление и своевременная поддержка позитивных активностей
 - эффективнее, чем формирование и навязывание инициатив сверху
 - эффективнее, чем сбор заявок на поддержку будущих активностей
 - является средством пассивного предупреждения негативных активностей
4. Возможные виды поддержки:
 - информационная, организационная, финансовая

Бизнес-логика процессов и потоки данных

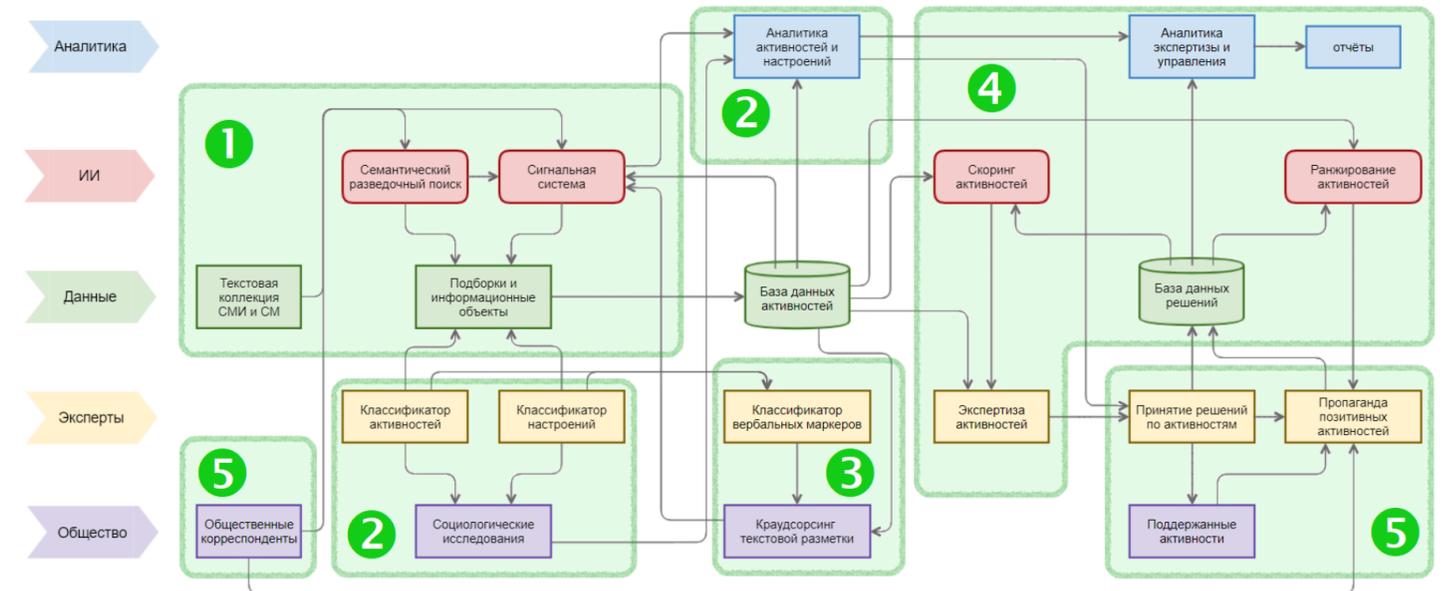


Бизнес-логика процессов и потоки данных



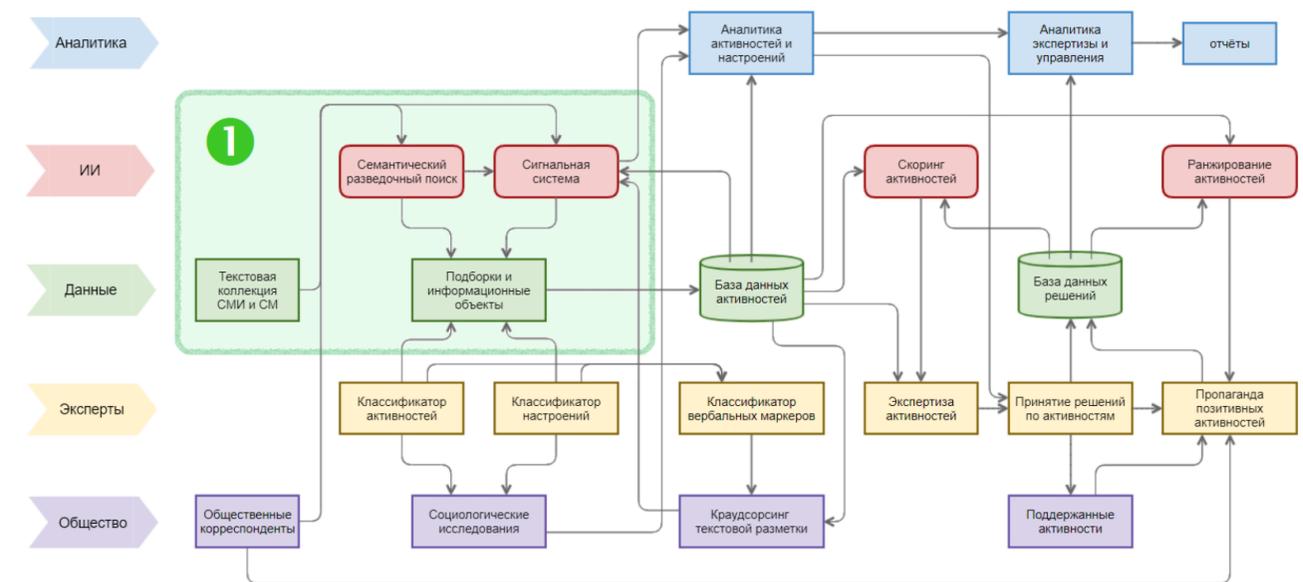
Пять взаимосвязанных контуров

- 1 Мониторинг СМИ и соцмедиа**
— формирование базы активностей
- 2 Социогуманитарная экспертиза**
— классификация активностей и общественных настроений
- 3 Лингвистическая экспертиза**
— выделение вербальных маркеров
- 4 Принятие решений**
— экспертиза и скоринг активностей
- 5 Исполнение решений и обратная связь**
— поддержка активностей, масштабирование и PR



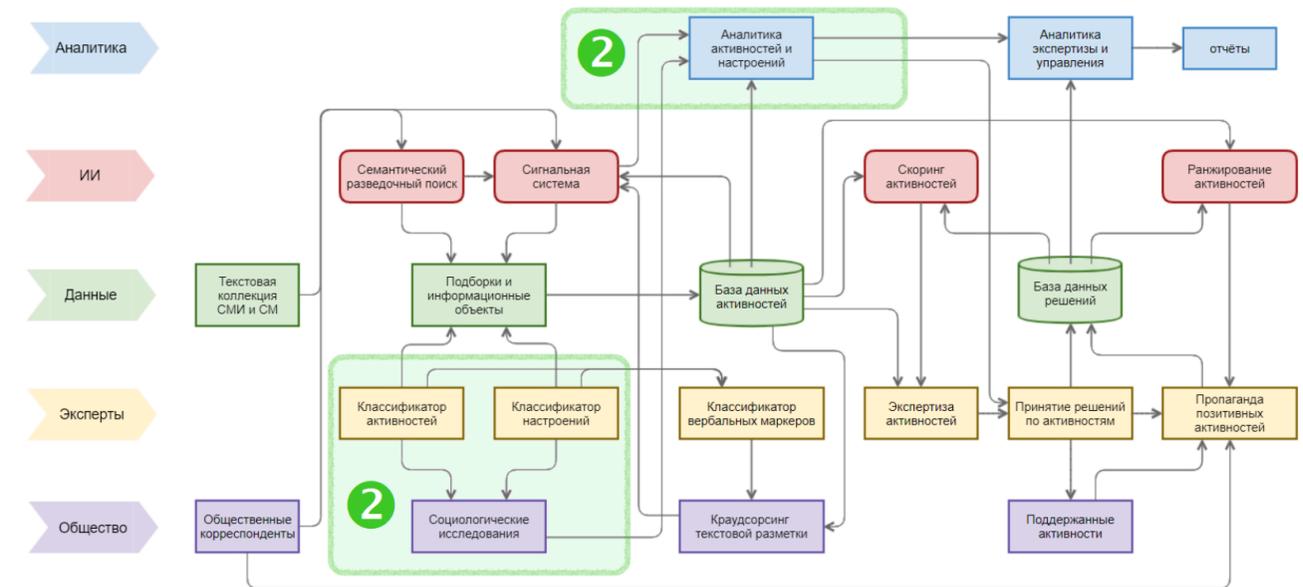
1 Мониторинг СМИ и соцмедиа

- Исходные данные:
 - новостные потоки СМИ
 - социальные медиа
 - общественные корреспонденты
- Семантический разведочный информационный поиск
 - поиск по смыслу, а не по ключевым словам
 - автоматизация рутинных операций поиска
 - формирование тематических подборок по активностям
- Сигнальная система
 - обнаружение новых активностей и общественных настроений
 - накопление данных о распространении активностей



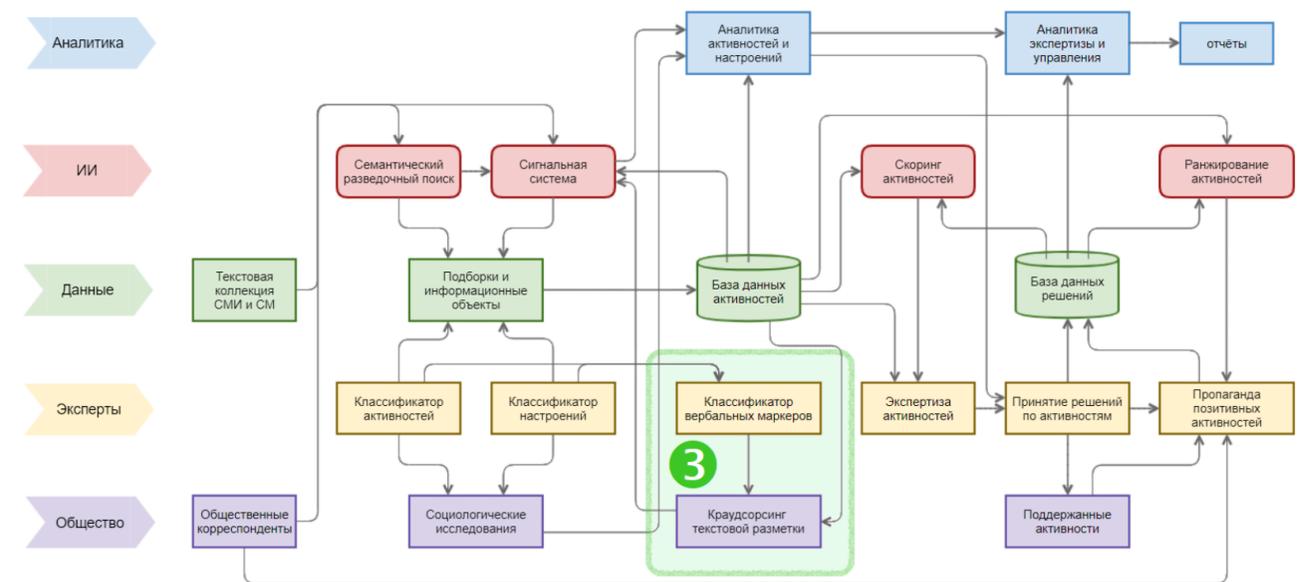
2 Социогуманитарная экспертиза

- **Классификатор активностей**
 - систематизация типов активностей
 - формирование «паспорта активности»
- **Классификатор общественных настроений и поляризации**
 - выявление ценностных ориентиров,
 - тональности общественных настроений
 - психологическая и психолингвистическая экспертиза
 - политологическая экспертиза
 - количественный и качественный анализ
- **Социологические исследования**
 - калибровка количественных оценок активностей и настроений
 - оценивание размеров соответствующих молодежных аудиторий и когортных групп



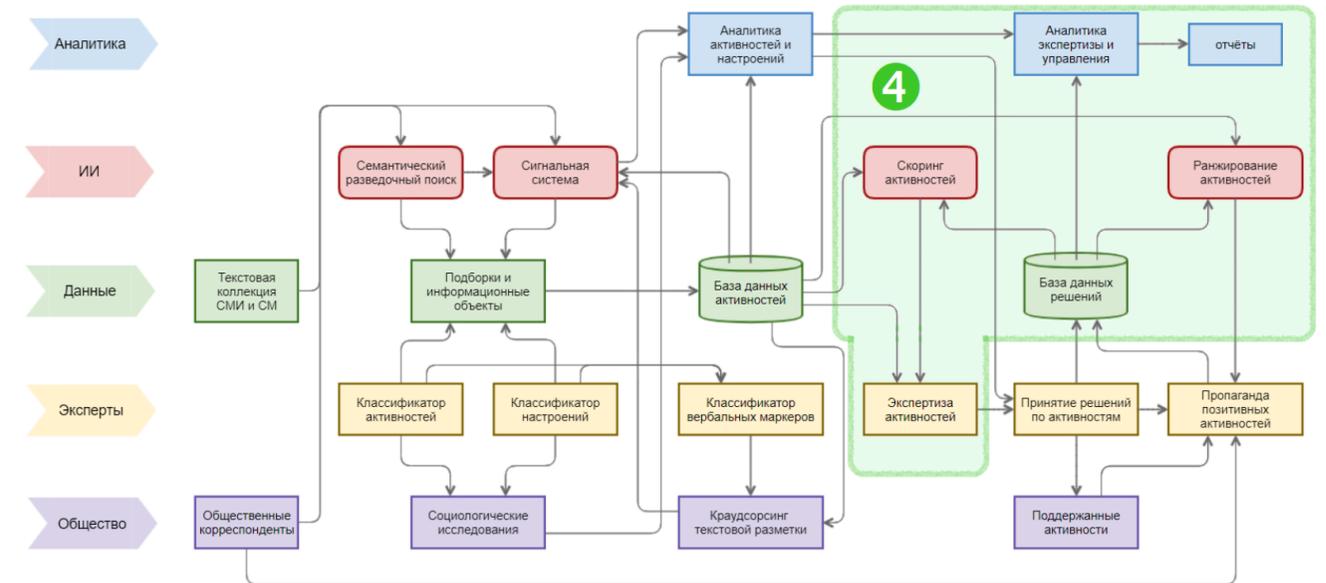
3 Лингвистическая экспертиза

- **Вербальные маркеры**
 - признаки для обучаемых моделей классификации активностей и настроений по текстовым материалам
 - привлечение экстралингвистических данных
- **Краудсорсинг**
 - методология разметки текстовых данных
 - связи между вербальными маркерами и классами активностей и настроений
- **Сигнальная система**
 - Размеченная выборка используется в контуре мониторинга для обучения моделей ИИ



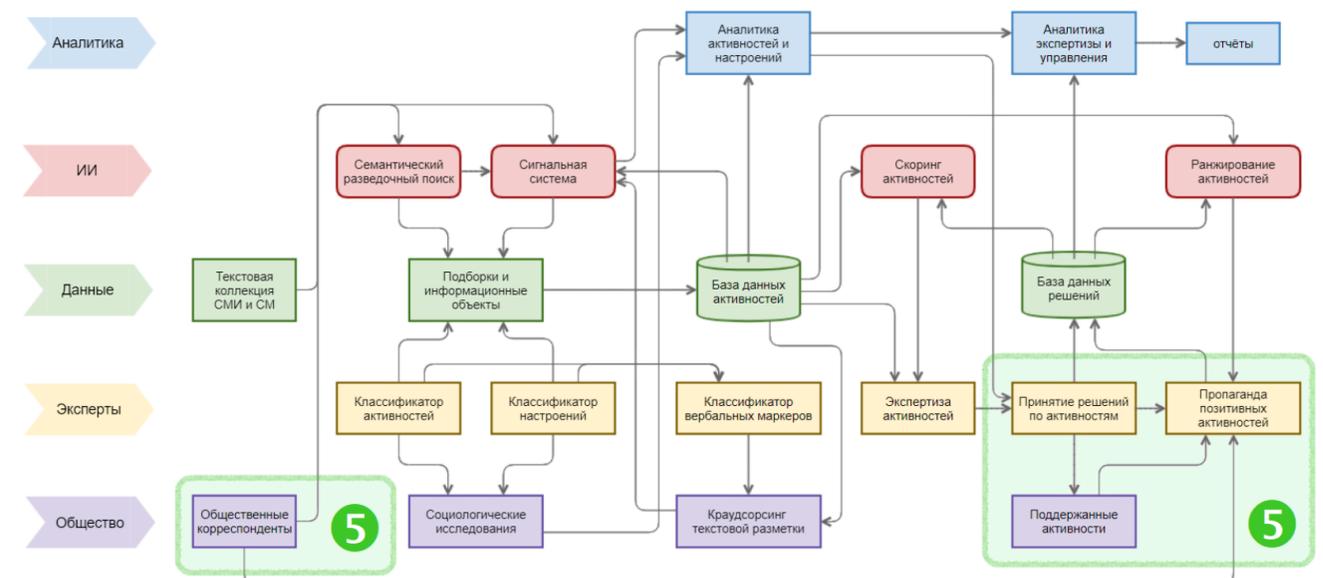
4 Принятие решений

- **Экспертиза активностей**
 - поддержка работы экспертов
- **Скоринг активностей**
 - автоматизация работы экспертов
 - генерация рекомендаций экспертам
 - ранжирование выявленных активностей
- **Принятие решений**
 - поддержка позитивных активностей
 - решения о видах и форме поддержки
 - пресечение негативных активностей
 - решения о каналах взаимодействия
 - формирование базы данных решений



5 Исполнение решений и обратная связь

- **Пиар позитивных активностей**
 - создание контента
 - организация обмена позитивным опытом
 - привлечение общественных корреспондентов
- **Исполнение решений**
 - информационная поддержка процессов исполнения решений
- **Обратная связь**
 - мониторинг случаев поддержки
 - количественные оценки эффективности управления



Университетский консорциум исследователей больших данных <http://opendata.university>



Университетский консорциум исследователей больших данных

Объединение научных и образовательных организаций, реализующих исследования и прикладные проекты в области сбора и анализа больших данных.

31 вуз

Участник Консорциума

7 центров

в университетах по всей России

>150 исследователей

в команде Консорциума



Индустриальные партнеры



Резюме

- *Цифровые технологии* (AI, ML, NLP, NLU) готовы для автоматизации решения широкого спектра социально значимых задач
- Имеется необходимая *междисциплинарная экспертиза*
- *Краудсорсинг* — один из магистральных способов формализации гуманитарных знаний (лингвистических, экстралингвистических, психологических, социологических, политологических и т.д.)

Спасибо за внимание!

Воронцов Константин Вячеславович
д.ф.-м.н., проф. РАН,
зав. лаб. Машинного Интеллекта МФТИ

[k.v.vorontsov @ phystech.edu](mailto:k.v.vorontsov@phystech.edu)

<http://www.MachineLearning.ru/wiki?title=User:Vokov>