summit

БОЛЬШИЕ ВЫЗОВЫ ДЛЯ ОБЩЕСТВА, ГОСУДАРСТВА И НАУКИ

Университет Сириус

Задачи и технологии понимания естественного языка: искусственный интеллект в помощь естественному

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

Перспективные задачи понимания естественного языка

Диалоговый интеллект

ответы на вопросы, автоматизация услуг и общения с клиентами

«Мастерская знаний»

поиск по смыслу, и что дальше делать с найденной информацией

Анти-фейк и анти-постправда

выявление потенциальных опасностей в дискурсе

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

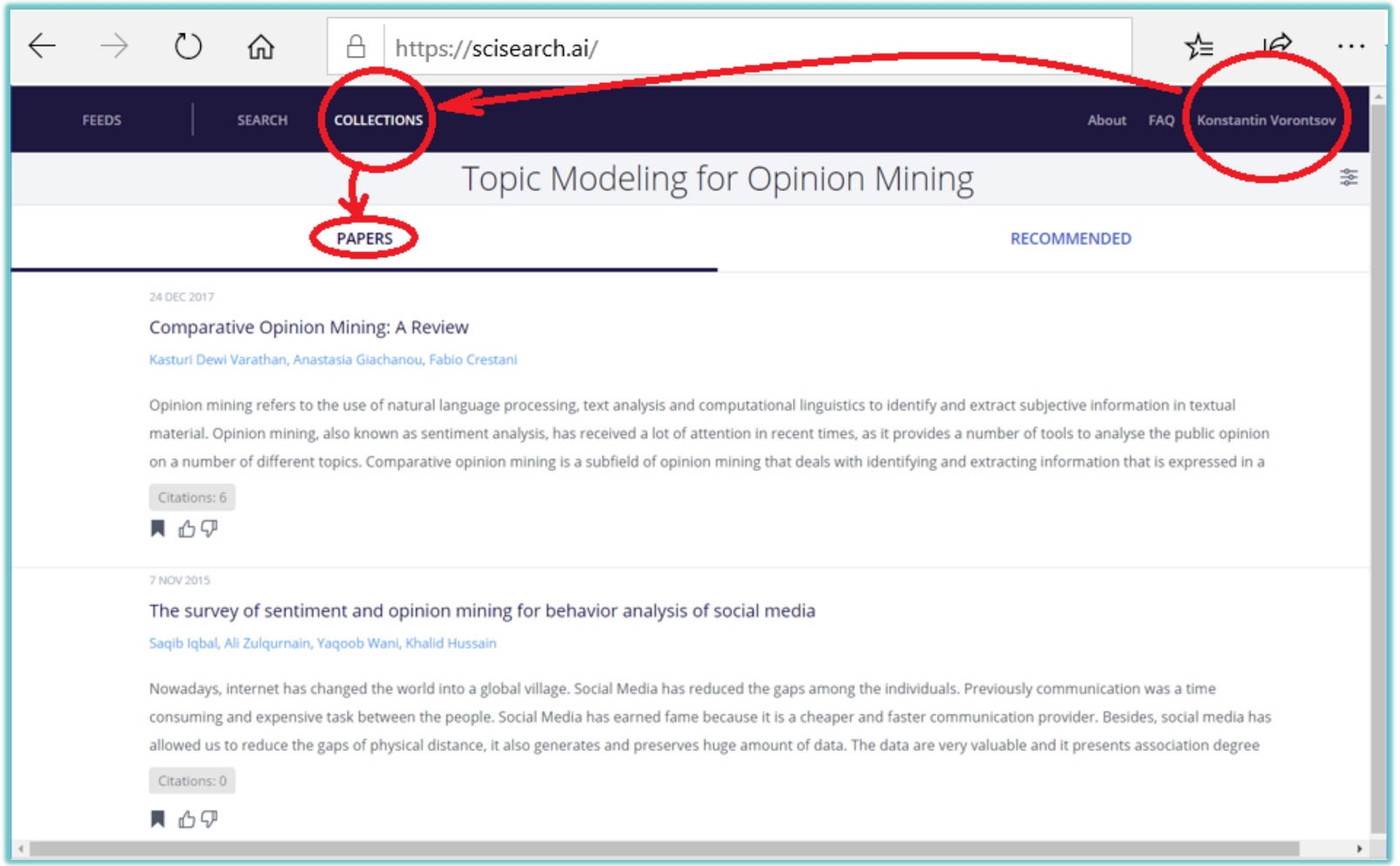
Аналитические функции:

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

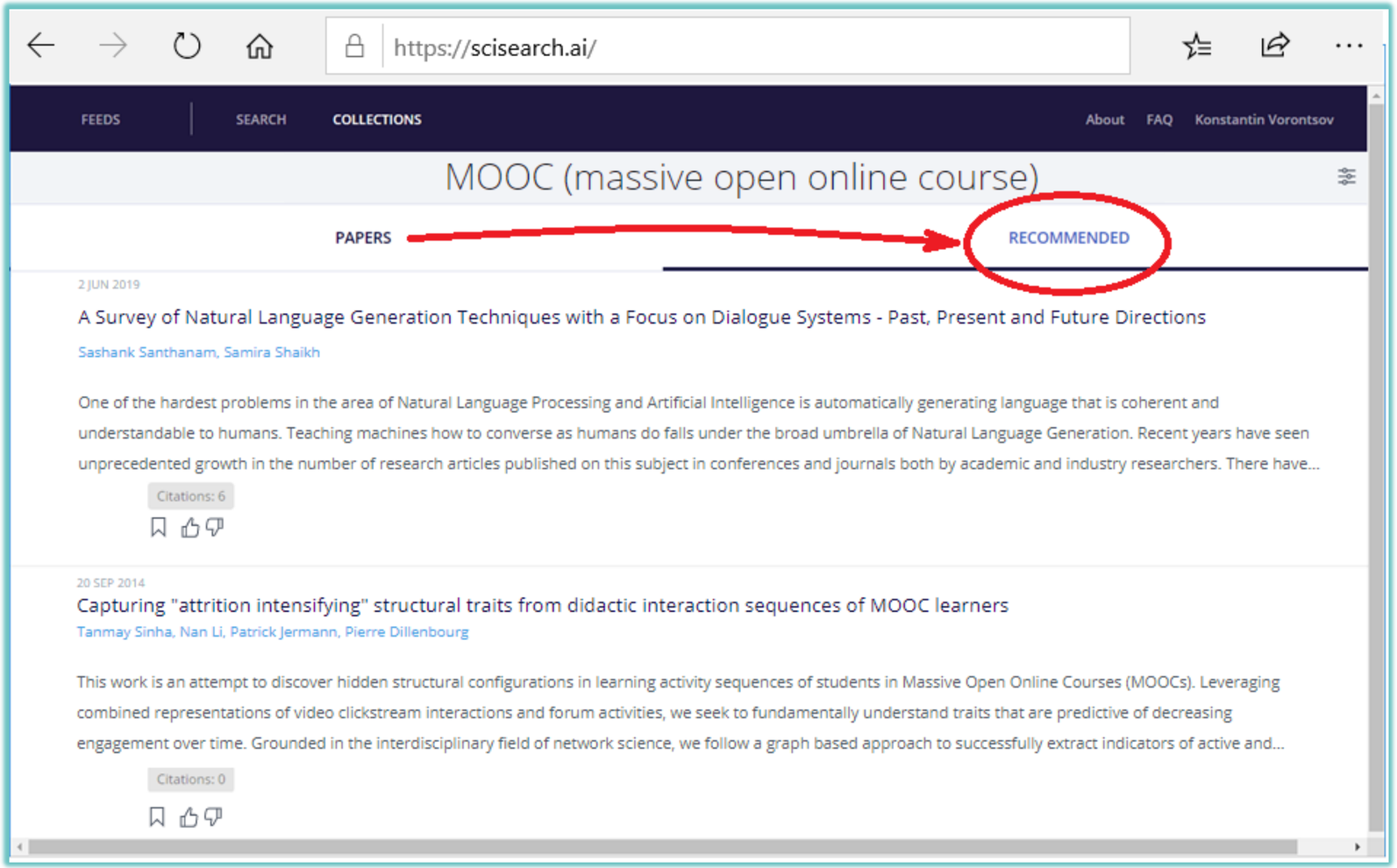
Коммуникативные функции:

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

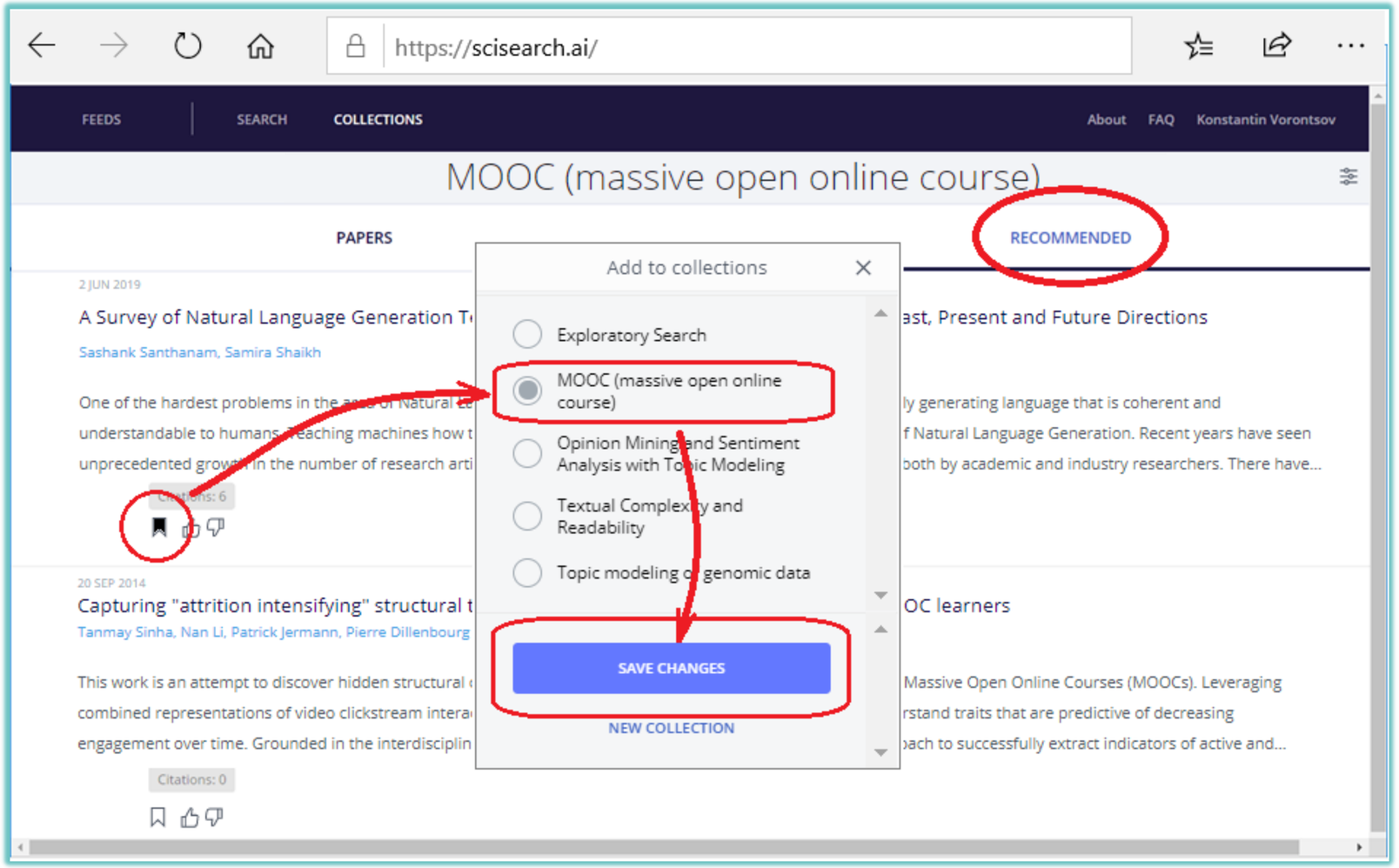
Поиск и рекомендации в SciSearch.ai



Поиск и рекомендации в SciSearch.ai



Поиск и рекомендации в SciSearch.ai



Технология тематического поиска BigARTM

Схема эксперимента:

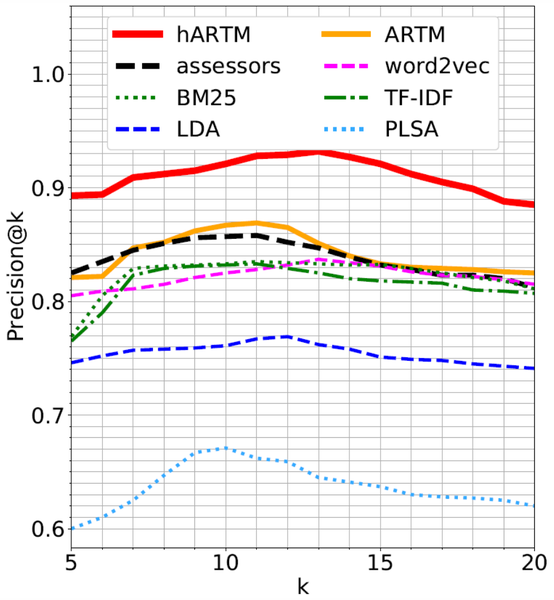
- длинные запросы (1 стр. A4)
- 100 запросов на коллекцию
- 3 асессора на каждый запрос
- от 10 до 60 минут на запрос
- разметка на Яндекс.Толока
- две коллекции техно-новостей:

(170K Russian docs)

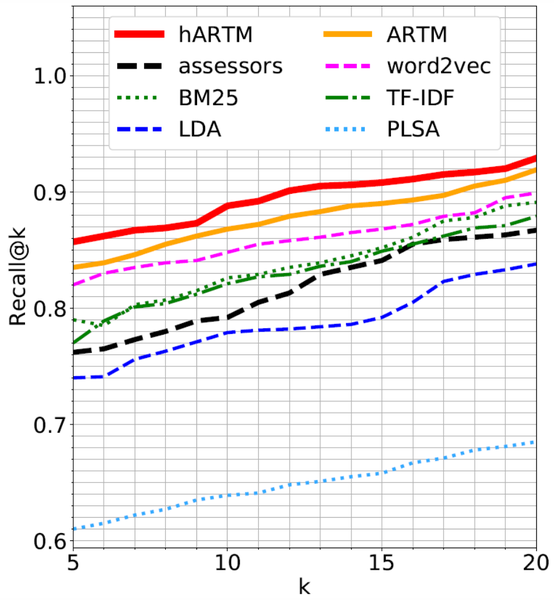
(750K English docs)

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:

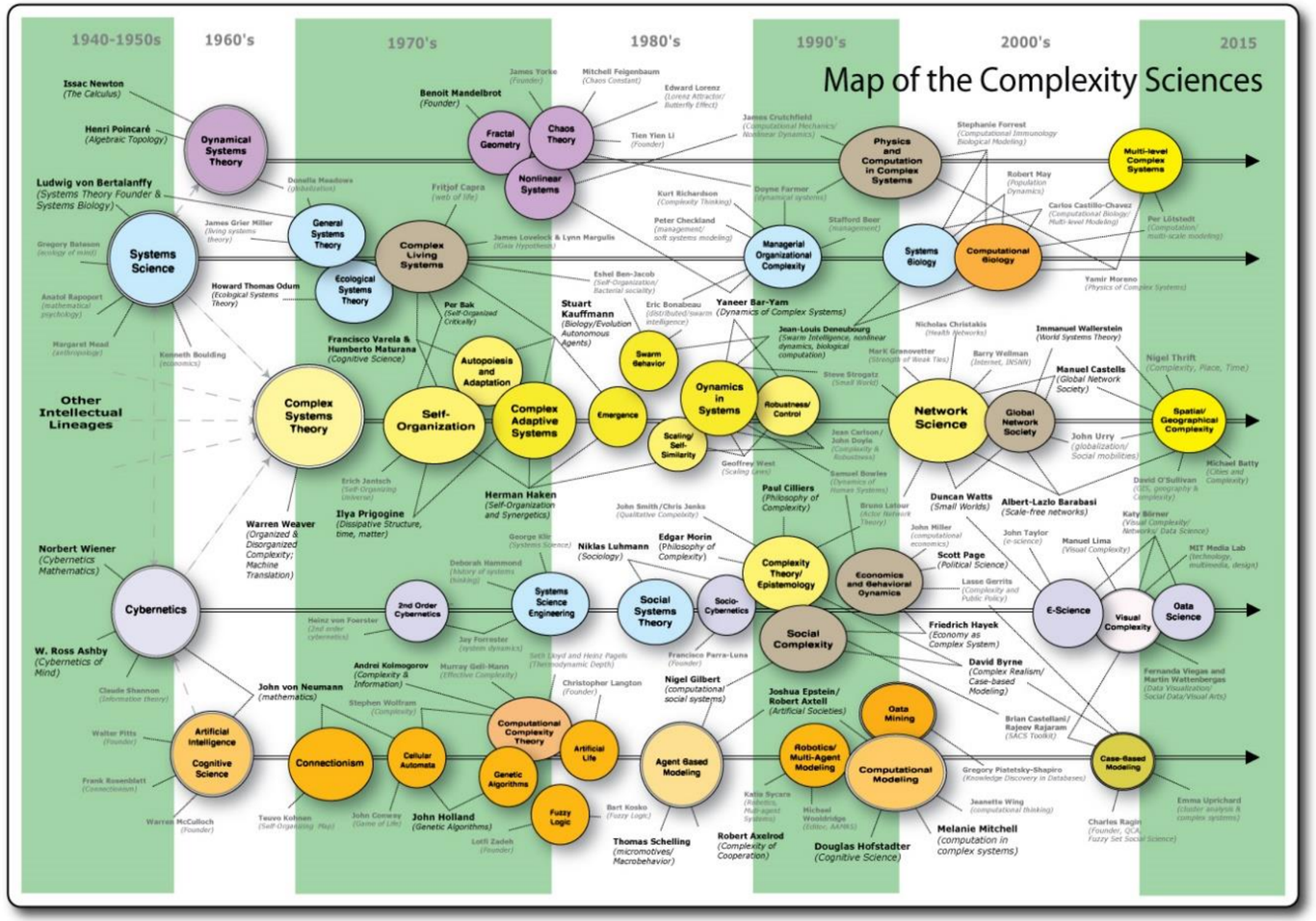
> Few contextual citation graphs are publicly available. The ACL Anthology Network (AAN) (Radev et al., 2009) is one such contextual citation graph built from the ACL Anthology corpus (Bird et al., 2008), consisting of 24.6K papers manually augmented with citation information. CiteSeer (Giles et al., 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)

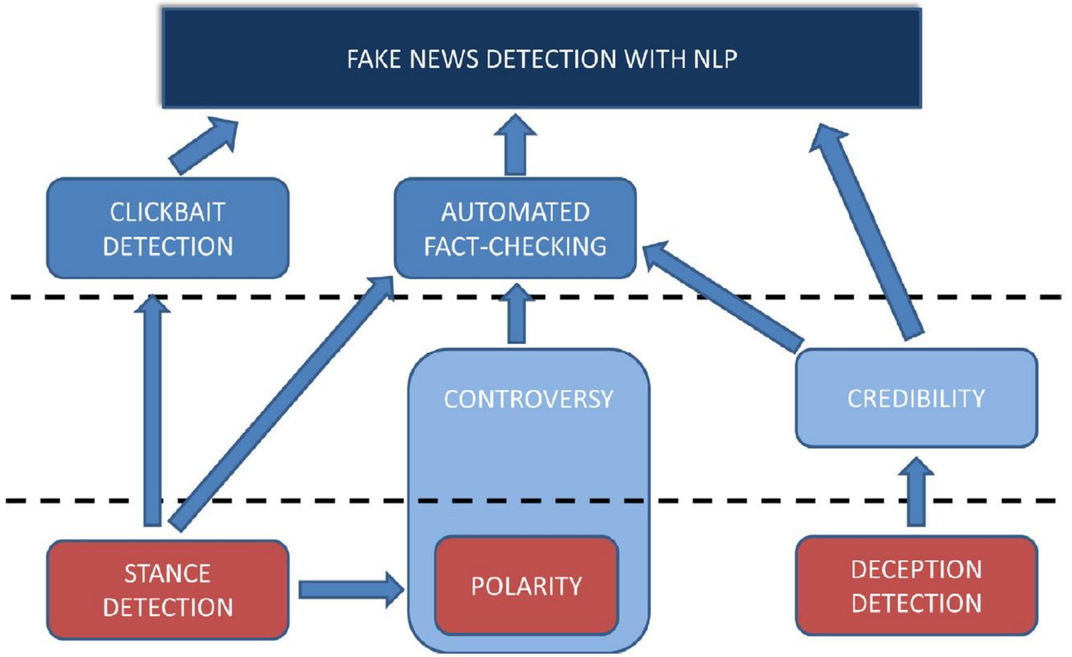
Осями на карте могут быть:

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



Область исследований «Fake News Detection»

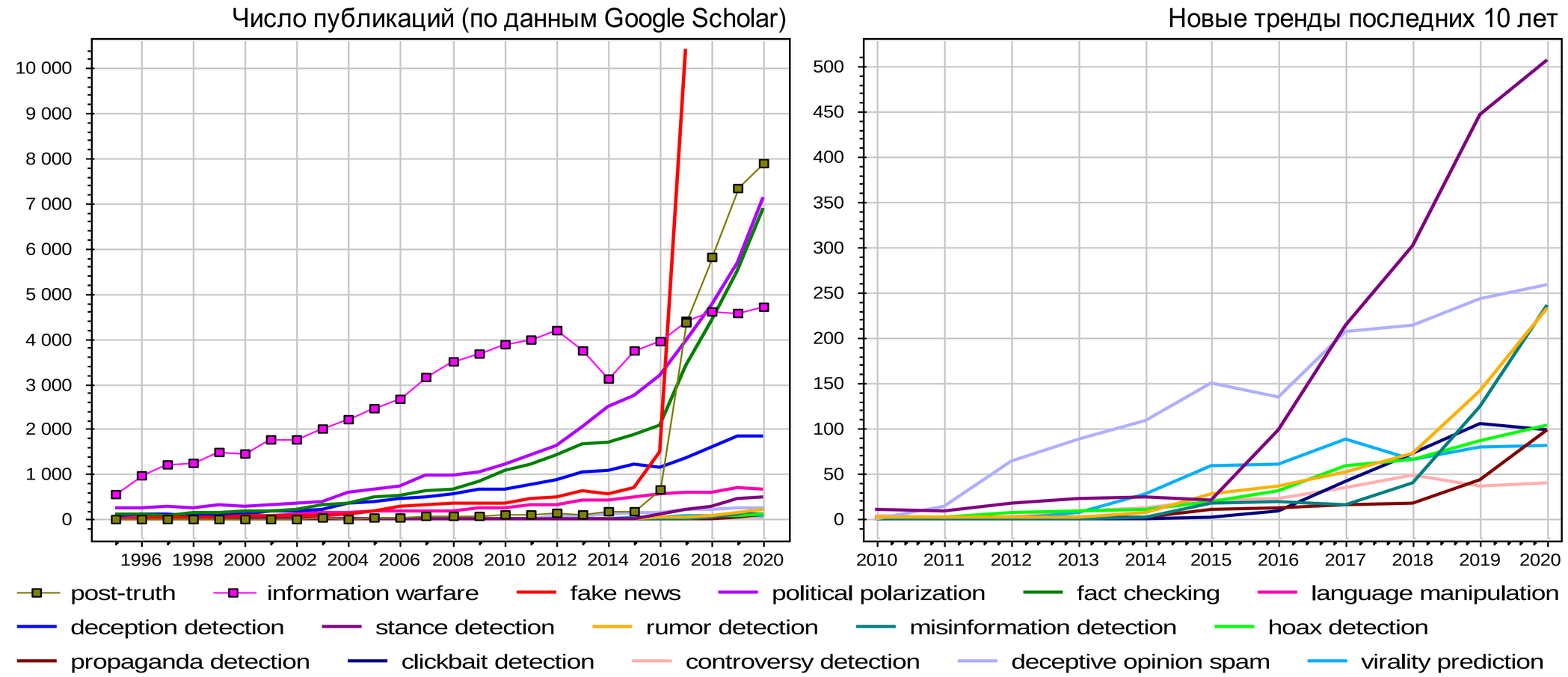
1. Deception Detection
   выявление обмана в тексте новости
2. Automated Fact-Checking
   автоматическая проверка фактов
3. Stance Detection
   выявление позиции за/против запроса (claim)
4. Controversy Detection
   выявление и кластеризация разногласий
5. Polarization Detection
   классификация позиций 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:



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 конструктов 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 конструктов 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

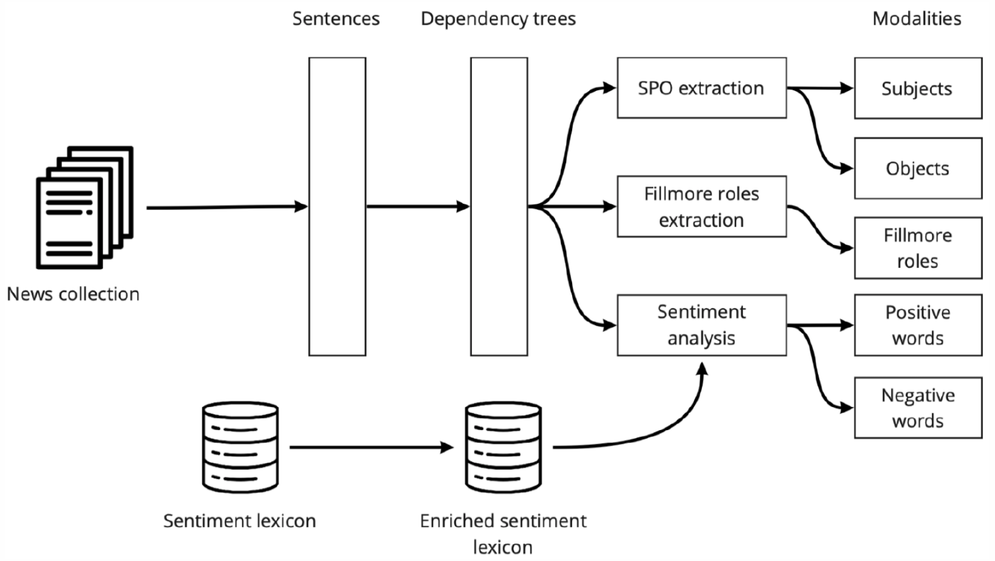
... Президент Петр Порошенко заявил, что Россия де-факто конфисковала украинские предприятия, которые находятся на неподконтрольной Киеву территории. Сегодня ДНР и ЛНР "национализировали" украинские предприятия ... При этом Кремль защитил конфискацию предприятий в ЛДНР ... Украина потребует расширить 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» — агенс 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	*Pr*	*Rec*	*F*1
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	*Pr*	*Rec*	*F*1
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.

Резюме

- Противостояние угрозам политики постправды – социально значимая задача, миссия и вызов для научно-технологического сообщества ML/NLP
- Проблематика *Fake News Detection* расширяется для выявления всех видов потенциально опасного дискурса (манипуляций, пропаганды, информационной войны)
- Эти задачи вполне решаемы современными средствами ML/NLP. Организационно это может быть проект с открытой концепцией и кодом
- Решение требует междисциплинарного подхода, объединения усилий AI-инженеров, лингвистов, психологов, политологов, журналистов

СПАСИБО!

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

k.v.vorontsov @ phystech.edu