Сравнительный анализ архивов научных журналов

Ф.В. Краснов (Газпромнефть НТЦ, 50+ научных статей, MBA, PMP, RHCE, к.т.н.),

М.Е. Шварцман (Библиотека имени Ленина, 90+ научных статей, к.т.н.),

А.В. Дементов (Национальный Электронно-Информационный Консорциум, elpub.ru)

СПб, 2019

Суть

Бизнес потребность:

Редакции журналов хотят понимать, что нужно изменить, чтобы войти в глобальные индексы цитирования (Scopus, WoS).

Наш подход:

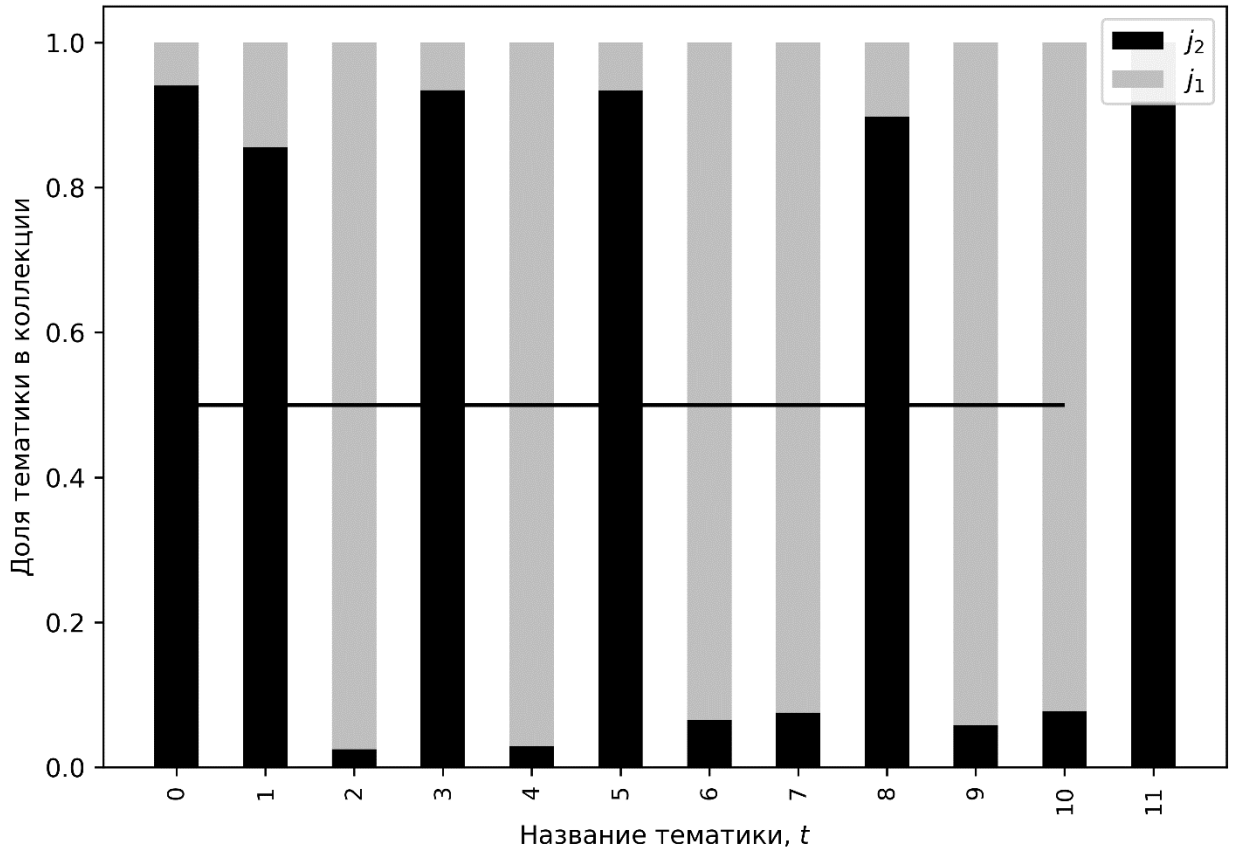
Бенчмаркинг архива журнала с мировыми лидерами с целью выявить отличия и выработать рекомендации по их устранению.

Результат (1)

- Онлайн сервис по бенчмаркингу архивов журналов,
- Полостью автоматизированный процесс,
- Комплексная методика на основе анализа текстов и мета-информации о научных статьях.

Сравнив архивы журналов, мы можем сказать редакции каких тематик, статей и авторов не хватает, чтобы удовлетворять требованиям Scopus и WoS.

Результат (2)



Коэффициент контентной аутентичности

Topic 0: hippocampal_subfield, subiculum, subfield, hippocampal_region, presubiculum, parasubiculum, hippocampal_tail,

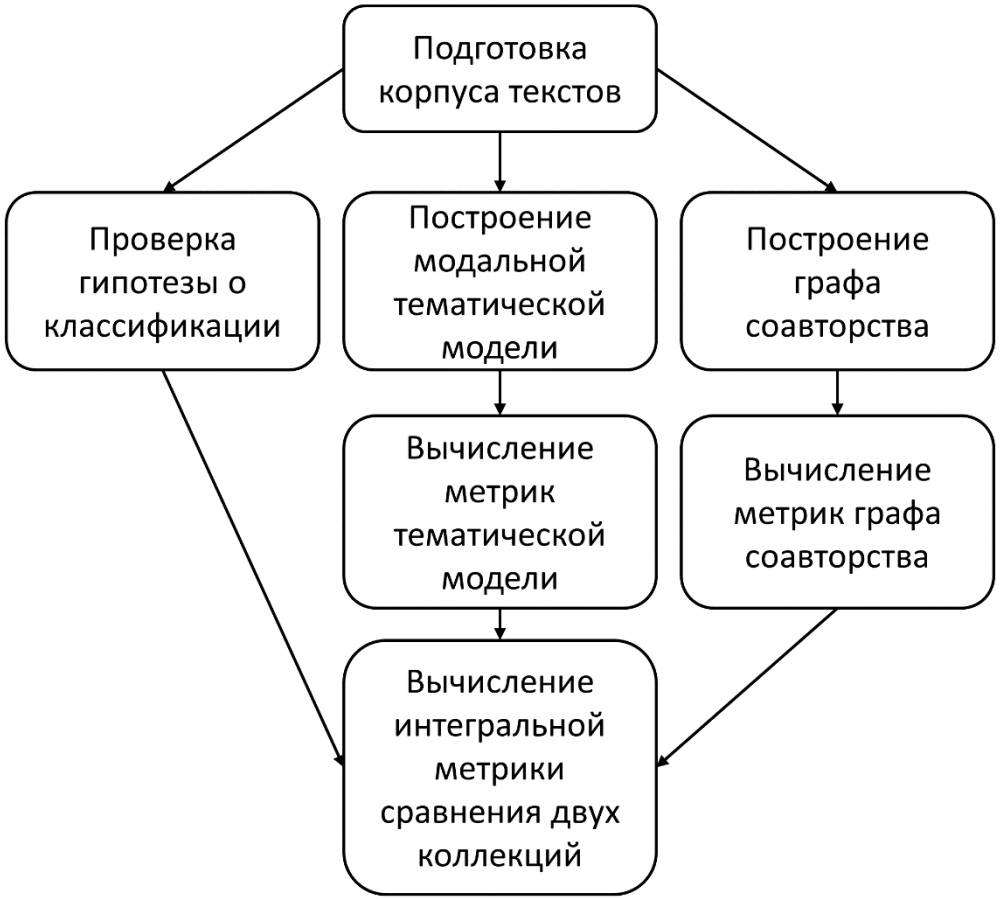
Topic 1: amygdala, fmri, ptsd, functional_magnetic_resonance_imaging, activation, emotion_regulation, ventromedial_prefrontal_cortex,

Topic 2: adhd, attention_deficit_hyperactivity_disorder, adhd_patient, intra_individual_variability, adolescent, erp, neurophysiological_datum,

Topic 3: multiple_sclerosis, flair, lesion, ms, segmentation, image, ms_patient,

Topic 4: priori_hypothesis, study_design, publication, subject_characteristic, medline, mr_spectroscopy, magnetization_transfer,

Результат (3)



Автоматизированный процесс:

1. Сбор данных;
2. Гипер-параметры;
3. Уменьшение «шума»;
4. Уровень автоматизации.

Методика Comparative Text Mining

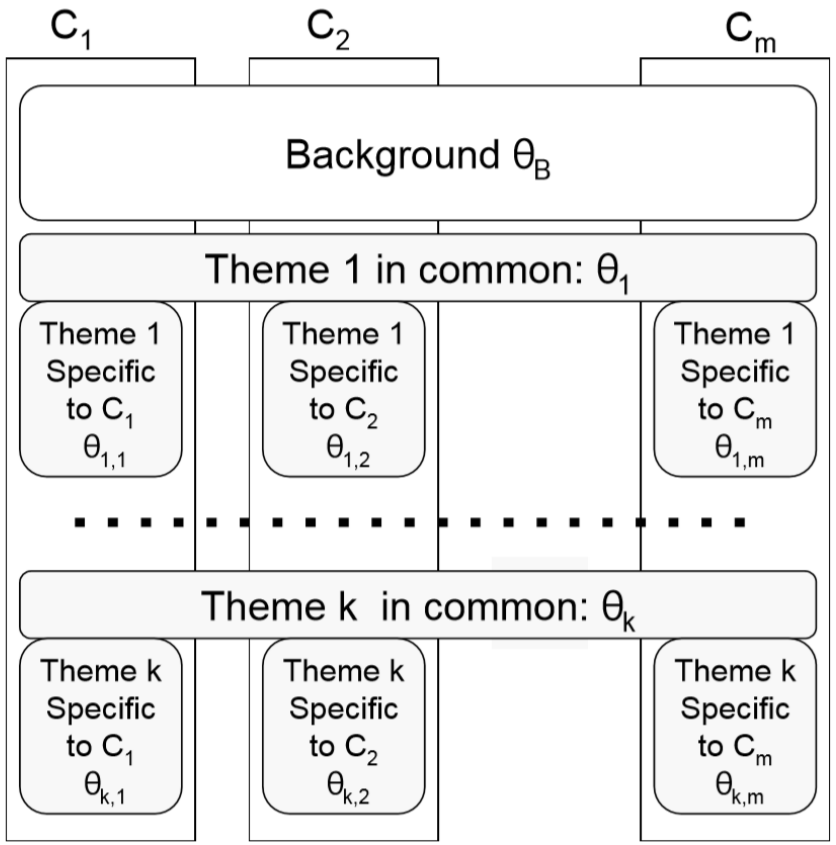
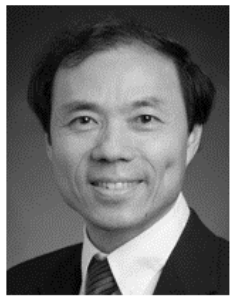


Figure 2: The Cross-Collection Mixture Model



ChengXiang ("Cheng") Zhai (Short Biography)

Donald Biggar Willett Professor in Engineering
Department of Computer Science
(Also affiliated with Carl R. Woese Institute for Genomic Biology, Statistics , and School of Information Sciences)
University of Illinois at Urbana-Champaign, 201 N. Goodwin Ave, Urbana, IL 61801
Office: 2116 Siebel Center (directions)
Phone: *+1-217-244-4943* **Fax:** *+1-217-265-6494*
Email: czhai AT illinois DOT edu
Assistant: Allison G. Mette (agk AT illinois DOT edu, *+1-217-300-6504*, 2106 SC)

Research Interests: My **general interests** are in developing all kinds of novel **Intelligent Information Systems** (e.g., **intelligent search engines**, **recommender systems**, **text analysis engines**, and **intelligent task assistants**) to help people manage and exploit large amounts of data (i.e., **"big data"**), especially **text data**. I am particularly interested in building such intelligent systems for **improving health, medical care, education, and accelerating scientific discovery**, and bulding them based on theoretically sound frameworks, models, and algorithms that are also effective empirically. My **current interests** include *Intelligent Information Retrieval*, *Text Data Mining*, *Applied Machine Learning*, *Optimization of Human-Computer Collaboration*, *Biomedical and Health Informatics*, and *Intelligent Educational Systems*. For a **more detailed view** of my research interests, you may want to take a look at the slides of some of my recent talks.

Research Group: TIMAN, DAIS

Publications & Recent Talks (Google Scholar Profile, DBLP Profile)

The Book Text Data Management: A Practical Introduction to Information Retrieval and Text Mining, ChengXiang Zhai and Sean Massung, ACM and Morgan & Claypool Publishers, July 2016. has now been translated into Chinese (see the Chinese version). The book is the main reference book for **two MOOCs on Coursera** that I'm teaching: (1) Text Retrieval and Search Engines ; (2) Text Mining and Analytics.

Научная новизна

- Мягкая кластеризация с помощью тематической модели
- Тематическая модель для аннотаций (коротких текстов)
- Увеличение отношения сигнал/шум с помощью фокусировки на частях речи (POST, Part-of-speech tagging)

Заключение

- Вывод на рынок нового аналитического продукта;
- От идеи до первой выручки 6 месяцев;
- Открытые программные каркасы (frameworks) можно продуктивно комбинировать, а не пытаться «допиливать»;
- Междисциплинарные команды экспертов высокоэффективны для инноваций.

Спасибо за внимание.

Федор Краснов: krasnov.fv (at) gazpromneft-ntc.ru

Михаил Шварцман: shvar (at) rsl.ru

Александр Дементов: dimentov (at) neicon.ru