

МОДЕЛИ АНАЛИЗА НАУЧНЫХ ТЕКСТОВ И СОПОСТАВЛЕНИЯ АКТОРОВ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ

Крейнес М. Г., Крейнес Е. М.
(ООО «БАЗИСНЫЕ ТЕХНОЛОГИИ», Москва)

Методология, модели и методы объективного анализа, оценивания и сопоставления конкретных научных результатов и акторов научной деятельности (отдельных ученых, коллективов, организаций, научных журналов, конференций, издателей научной литературы и т. п.) сегодня является предметом научной и экспертной дискуссии. Распространенные в настоящее время подходы имеют существенные недостатки, осознаваемые всеми заинтересованными сторонами. Широко тиражируемые библиометрические и наукометрические методы подвержены манипулированию (со стороны бизнеса и со стороны субъектов научной деятельности) и не всегда научно обоснованы. Экспертные методы дорогостоящи, не свободны от вкусовщины и также легко поддаются манипуляциям. Трюизмом стало утверждение о необходимости формального объективного анализа содержания и автоматической тематической категоризации научных текстов для корректного применения методов библиометрии и наукометрии. Несмотря на явно сформулированный запрос, серьезные достижения в данной области не известны.

Задачу построения объективных оценок научных текстов и акторов научной деятельности удалось поставить и решить с применением разработанных нами вычислительных моделей содержания (семантики) текстов на естественных языках, моделей оценки семантической близости текстов, моделей выявления тематик текстовых коллекций, моделей оценки их тематической близости (обзор соответствующих работ [1]), а также моделей вычисления характеристик содержания научного текста относительно текстовой коллекции [2]. Разработаны учитывающие данные библиометрии/наукометрии и результаты вычислительного моделирования текстов и текстовых коллекций - формаль-

ные методы решения задач оценки качества научных текстов и акторов научной деятельности:

- поиск содержательно схожих текстов и выявление среди них сопоставимых научных текстов (коллекции текстов, адекватных для сравнения с анализируемым текстом),

- объективное оценивание текста по характеристикам содержания и библиометрическим показателям анализируемого и сопоставимых с ним текстов, выявление содержательно необоснованного («дружественного») цитирования по формальным расчетным критериям,

- оценка актора научной деятельности по объективным оценкам текстов, включенных в коллекцию, ассоциированную с актором, и выявление акторов научной деятельности, злоупотребляющих дружественным цитированием,

- поиск акторов, порождающих тематически близкие текстовые коллекции, по формируемым вычислительно моделям тематики текстовых коллекций, ассоциированных с акторами научной деятельности,

- объективное оценивание актора научной деятельности по результатам сравнения с сопоставимыми акторами.

Результаты вычислительных экспериментов с миллионом русскоязычных научных статей (2009 – 2015 г.г., elibrary.ru) и 180 тысячами англоязычных статей (журнал Science с начала издания журнала до 1996 г.) (данные предоставлены НЭИКОН) показали практическую применимость предложенных моделей.

Литература

1. Крейнес М.Г. *Модели текстов и текстовых коллекций для поиска и анализа информации*// Труды Московского физико-технического института (государственного университета). – 2017. – т. 9. – № 3 (35). – С. 132-142.
2. Крейнес М.Г. *Методы вычислительного анализа моделей семантики для оценки качества научных текстов* // Известия РАН. Теория и системы управления. – 2013. – № 2. – С. 64-75.

