

От просопографии университетской профессуры до цифрового следа философского парохода: проблемы и результаты работы неформальной группы по применению количественных методов в истории науки и социологии научного знания

Алексей Куприянов

Приблизительно с 2007 г. в НИУ ВШЭ (СПб) работает неформальная группа вынужденно непостоянного состава (значительную часть ее составляют студенты бакалаврской программы), занимающаяся проектами, близкими к Digital humanities / Computational social science. Организационно она связана с Научно-исследовательским семинаром по социальным исследованиям науки и технологий под руководством М. Р. Демина и автора этих строк. Проекты, при всем их разнообразии, связаны с двумя системами баз данных, в одной из которых концентрируются сведения о научных и философских журналах, преимущественно СССР и Российской Империи, в другой -- о профессорах и преподавателях университетов Российской Империи. Предварительные результаты работы, помимо курсовых работ и бакалаврских диссертаций нашли отражение в ряде публикаций и выступлений (см. Куприянов 2013, Иванова 2014, Фотиади 2014, Kouprianov 2014). В настоящем сообщении будет дан обзор основных результатов работы и проблем как концептуального, так и организационного плана, с которыми мы столкнулись.

Для реализации наших проектов нам требовались (1) содержательно богатые, полные, большие массивы информации, (2) модели, которые можно тестировать на этих массивах.

Под содержательно богатыми я подразумеваю данные, в которых каждая из описываемых сущностей охарактеризована по многим атрибутам. Под полнотой я подразумеваю то, что сколь возможно большее количество описываемых сущностей охарактеризовано по всем атрибутам. Размеры используемых нами массивов информации различны, однако, как правило, они, по меркам Digital Humanities / Computational Social Sciences, не очень велики (несколько тысяч профессорских биографий, около двух тысяч наименований журналов, несколько десятков тысяч библиографических описаний статей из научных и философских журналов, несколько сотен оцифрованных полных текстов и т. д.). Причины этого отчасти кроются в природе материала (в конце-концов, профессоров университетов дореволюционной России уже никогда не станет больше, чем их уже было), отчасти, увы, в ограниченности ресурсов (проект осуществляется в инициативном порядке, без какой-либо поддержки). Под моделями я подразумеваю то, что позволяет выстроить гипотезы о необходимых или, хотя бы, вероятностных связях между разными атрибутами.

С какими проблемами нам пришлось столкнуться? Главная методологическая -- выделение сущностей и пространств отношений между ними (это общие проблемы организации всех реляционных баз данных). Главная практическая -- размеры и полнота массивов.

Основной путь преодоления этих проблем -- коллаборативные проекты. Обсуждение с коллегами в несколько раундов позволяет еще на этапе разработки и тестирования структуры базы данных избежать неверных шагов, которые в перспективе могут создать серьезные принципиальные ограничения для дальнейшего роста базы или анализа. Кроме того, поскольку мы работаем, в основном, с материалами, которые невозможно собрать в автоматическом режиме (ряд подпроектов держится на размещенных в открытом доступе в Интернете материалах, которые можно собрать и свести в базу при помощи роботов, но это

не основной наш "профиль"), много времени приходится уделять ручному вводу информации с бумажных носителей в базы и оцифровке полнотекстовых первичных источников. Когда в проект удастся вовлечь хотя бы трех-четыре человека, дело идет быстрее.

В коллаборативных проектах однако на первый план выходят организационные и этические вопросы. Например, необходимость "стандартизации" наблюдателя -- того, кто преобразует неструктурированную или частично структурированную информацию в формат базы данных и создание интуитивно-понятного формата ввода данных в базу. Недостаточная "стандартизация" и переусложненный формат ввода приводят к возникновению избыточного "шума" в данных. Это только усугубляет проблемы, связанные, например, с вариативностью написания имен и разночтений в датах, имеющей место в первичных источниках нашего типа (обращение к различным источникам необходимо, поскольку ни один из них, взятый по отдельности, не содержит обычно необходимую нам информацию по большинству атрибутов в полном объеме). Довольно сложно организована и моральная экономика обращения с общим ресурсом (как авторизовать вклад наблюдателей, делить авторство в публикациях и т. д.). Нельзя сказать, чтобы эти проблемы во всех случаях удавалось решать удачно, но постепенно складываются представления о том, как должна выглядеть система частично эксплицируемых правил, которая помогла бы минимизировать конфликты.

К наиболее значимым результатам работы группы, помимо создания системы баз данных, можно отнести (1) разработку подходов к анализу библиометрических кризисов и демонстрацию различного характера кризисов, которые претерпевала система научной периодики СССР (Куприянов 2013, Kourgiarov 2014), (2) разработку формального анализа структуры авторских коллективов журналов и картирование поля русскоязычных философских журналов первой половины XX века (Фотиади 2014), (3) разработку подходов к созданию картины долговременной динамики кадрового состава университетов дореволюционной России (Иванова 2014). Из неопубликованного весьма перспективными представляются мне результаты проектов, связанных с контент-анализом изображений в научных журналах (Валеева 2010), историей становления научных дисциплин в университетах (Копова 2013) и анализом корпуса текстов некрологов советских ученых.

Список литературы

- Валеева Д. Р.* Различные способы конструирования научной убедительности: количественное исследование способов визуальной презентации данных в биологических и сельскохозяйственных журналах СССР, 1919-1965 гг. Выпускная квалификационная работа бакалавра социологии / Научный руководитель А. В. Куприянов. СПб: Высшая школа экономики, 2010.
- Иванова Е. В.* Что может дать единая база данных по профессуре дореволюционной России исследователям академической мобильности? // Социология в действии — 2014. Избранные материалы VI социологической межвузовской конференции студентов и аспирантов / отв. ред. М. Р. Демин. СПб.: Отдел оперативной полиграфии НИУ ВШЭ — Санкт-Петербург, 2014. С. 131-141.
- Копова В. М.* Формирование научных дисциплин во второй половине XVIII – начале XIX вв.: случай Императорского Московского университета. Выпускная квалификационная работа бакалавра социологии / Научный руководитель А. В. Куприянов. СПб: Высшая школа экономики, 2013.
- Куприянов А. В.* Реструктуризация и общая депрессия: предварительные замечания о природе библиометрических кризисов в истории Советской науки // Социология науки и технологий. 2013. Т. 4. № 4. С. 80-98.
- Фотиади М. Ф.* Философия революционной эпохи: наукометрический анализ русской философской периодики первой половины XX века // Социология в действии — 2014.

Избранные материалы VI социологической межвузовской конференции студентов и аспирантов / отв. ред. М. Р. Демин. СПб.: Отдел оперативной полиграфии НИУ ВШЭ — Санкт-Петербург, 2014. С. 120-131.

Kouprianov A. V. Beyond the Humanities: A Comparison of two Bibliometric Crises in the Domain of Soviet Biological Periodicals (1917–1950) // Russian Journal of Communication. 2014. Vol. 6. No. 1. P. 52–66.

--

Куприянов А. В., канд. биол. наук, департамент социологии, Национальный исследовательский университет -- Высшая школа экономики (Санкт-Петербург).

ул. Союза Печатников, д. 16, Санкт-Петербург, Россия, 190008

E-mail: alexei.kouprianov@gmail.com

<http://www.hse.ru/org/persons/202735>