

УИС РОССИЯ: БАЗА ГЕОГРАФИЧЕСКИХ ЗНАНИЙ ПО РФ ДЛЯ ПОДДЕРЖКИ РЕГИОНАЛЬНЫХ ИССЛЕДОВАНИЙ

Т.Н. Юдина, В.В. Гребенюков

*Московский государственный университет имени М.В. Ломоносова
Москва*

В 2017 году в рамках работ по развитию Университетской информационной системы РОССИЯ разработана База географических знаний по РФ. База доступна на сайте УИС РОССИЯ (<http://uisrussia.msu.ru>) План проекта на 2018 год — разработка технологии для анализа аналитических публикаций и статей СМИ с использованием Каталога географических названий — выделения географических объектов и приписывания каждой публикации и статьи к соответствующему муниципалитету и региону.

Географические знания — необходимый компонент информационных систем для региональных социально-экономических исследований: системный анализ ситуации в регионах и на местах предполагает обработку и содержательную интеграцию широкого и постоянно растущего круга источников нескольких типов, характеризующих разные аспекты жизни граждан и положение дел в каждом регионе. Корректность обработки документов и их организации в систему определяют информационный потенциал и аналитическую ценность системы и напрямую зависят от способности программного обеспечения выявлять географическую принадлежность события (обрабатывать и атрибутировать географические объекты) и приписывать каждый объект к соответствующей территории, т.е. организовывать входящий поток документов по регионам.

Основу информационной инфраструктуры для региональных исследований составляют статистические показатели и аналитические публикации — издания органов власти, научных центров, статьи и сообщения СМИ. Статистические данные из сборника «Регионы России» и почти 20 других сборников Росстата загружаются в базу данных. Географическая информация в аналитических документах представлена в виде текстов, обработка возможна только по упомянутому географическому названию. Анализ документов с использованием географической базы знаний позволит организовать информационный поток по регионам с учетом упоминаемых в документах географических единиц разных типов (природные объекты; республики, края, области, города федерального значения, автономная область, автономные округа; города и другие поселения, районы, железнодорожные станции и другие объекты). Такая организация документов — интеграция статистических показателей и документов умножит аналитический потенциал исследований по регионам РФ для всех гуманитарных дисциплин. При этом системные исследования будут возможны на уровне муниципалитетов, городов, поселков и деревень — уровне реальной жизни людей. Мониторинг природных и экологических процессов в регионах и на местах стимулирует исследования и разработку проектов рационального природопользования и охраны природы, а также изучения последствий неблагоприятных событий для здоровья населения на конкретной территории, выявления территорий и факторов риска. Могут быть также определены наиболее благоприятные и благополучные районы в каждом регионе.

База географических знаний может быть построена только на корректном источнике географических данных. В нашем случае — это Государственный Каталог географических названий (ГКГН). Каталог — официальный первоисточник географических названий РФ, включает полный список географических названий основных типов объектов и поддерживается в актуальной версии ФГБУ «Центр геодезии, картографии и инфраструктуры пространственных данных». Данные в Каталоге представлены в виде специальных системных таблиц. Каталог стал доступен на открытом сайте Росреестра с 2016 года в форматах pdf и csv [1]

Второй источник данных — Общероссийский классификатор объектов административно-территориального деления (ОКАТО). Поддерживается Росстатом [2].

Формирование базы географических названий по РФ и представление в открытом доступе на сайте УИС РОССИЯ — результат первого этапа проекта. На втором этапе проекта разрабатывается технология для обработки аналитических публикаций и статей СМИ с использованием Каталога географических названий — выделения географических объектов и приписывания каждой публикации и статьи к соответствующему муниципалитету и региону. Коллекции аналитических публикаций и статей СМИ по социально-экономической тематике поддерживаются в составе Университетской информационной системы РОССИЯ и включают около 30 источников — доклады государственных организаций (Счетная палата, Центральный банк, Министерство экономического развития и другие), публикации аналитических центров (Экономическая экспертная группа, Центр макроэкономического анализа и краткосрочного прогнозирования, Леонтьевский центр, Всемирный банк), научные издания (журналы экономического

факультета МГУ имени М.В.Ломоносова, НИУ ВШЭ, Института социологии РАН), СМИ (газета «Коммерсант и приложения, издательство РИА «Новости»).

Первая версия программы анализа текстов показала, что обработка сообщений СМИ проще в силу небольшой длины сообщения. Пример сообщения РИА Новости от 21 сентября 2017 — «Министерство природы составило список 20 городов с наибольшим уровнем загрязнения атмосферы, со ссылкой на государственный доклад ведомства "О состоянии и об охране окружающей среды России". В топ-20 вошли Биробиджан, Благовещенск, Братск, Зима, Красноярск, Кызыл, Лесосибирск, Магнитогорск, Минусинск, Новокузнецк, Норильск, Петровск-Забайкальский, Селенгинск, Улан-Удэ, Усолье-Сибирское, Чегдомын, Черемхово, Черногорск, Чита, Шелехов.

В большинстве городов Дальневосточного, Сибирского и Уральского федеральных округов уровень концентрации загрязняющих веществ превышает предельно допустимый. Лидерами по загрязнению воздуха оказались Красноярский и Алтайский края, Кемеровская, Иркутская и Новосибирская области. В четырех последних в 2016 году показатели выросли по сравнению с прошлыми годами» [3].

Вторая часть сообщения будет приписана к каждому из упоминаемых 8 регионов.

Обработка первой части сообщения требует анализа и уточнения. Так, Биробиджан встречается в Каталоге один раз, и сообщение будет приписано к Биробиджану и Еврейской автономной области. Братск встречается 3 раза — как город областного подчинения, международный аэропорт и ж-д станция. Сообщение будет приписано к городу областного подчинения и к Иркутской области. Зима встречается 11 раз в шести регионах, но это реки и ручьи в Магаданской, Сахалинской областях и Республике Саха, только один город областного подчинения и ж-д станция в Иркутской области. Красноярск — город краевого подчинения в Красноярском крае и село в Оренбургской области. Кызыл, Лесосибирск, Новокузнецк, Чегдомын, Черемхово, Черногорск, Шелехов встречаются либо один раз, либо могут быть легко выделены как город или рабочий поселок (Чегдомын). Благовещенск встречается 5 раз (город областного подчинения в Амурской области, город республиканского подчинения в Республике Башкортостан, международный аэропорт в Хабаровском крае, ручей в Амурской области, остановочный пункт в Архангельской области). Алгоритм определения географической принадлежности такой — первым шагом исключаются остановочный пункт, ручей, аэропорт. В первой версии программы обработки будет выбран город областного подчинения в Амурской области.

Первая часть сообщения будет приписана к каждому из 20 городов и к соответствующему региону.

Обработка аналитических публикаций — более сложная процедура. Возьмем текст статьи в Вестнике МГУ. Серия 6. Экономика — «Несмотря на наличие огромных запасов энергетических ресурсов, Республика Саха (Якутия) находится практически в полной зависимости от завозимых извне дорогостоящих нефтепродуктов. В результате этого в общем объеме затрат на энергопотребление более 70% стоимости потребления топливно-энергетических ресурсов принадлежит привозным нефтепродуктам. Например, в арктических улусах затраты на завоз топлива превышают стоимость самого топлива в 3–4 раза» [4]. Сообщение приписано к Республике Саха (Якутия).

В той же статье «Одним из решений проблемы энергообеспечения может стать производство пеллет. Пеллеты — биотопливо, производится из отходов деревообрабатывающей промышленности и сельского хозяйства. В Европе такое биотопливо широко используют в содержании частных домов. Материал отличается высокой энергоэффективностью, — экологичность, безотходность и легкость в транспортировке. Красноярский край, Брянская область, Вологодская область, Нижегородская область, Республика Коми, Республика Бурятия вышли в лидеры производства пеллет в России. Сообщение будет приписано к каждому из 6 регионов [4, с. 32].

Другая статья в этом же журнале анализирует экологическую обстановку в регионах страны, выделяет неблагоприятные регионы и города. «Неблагоприятная экологическая обстановка оказывает негативное влияние на здоровье населения. Смертность населения от новообразований устойчиво ассоциирована с загрязнением атмосферного воздуха бензапиреном, формальдегидом в 12 регионах. К территориям, подверженным наибольшим рискам, относятся Челябинская область, Забайкальский край, Курганская область, Приморский край, Нижегородская, Свердловская, Кемеровская области, г. Санкт-Петербург, Республика Бурятия, Волгоградская область, г. Москва, Вологодская область.

Наиболее высокий риск повышения заболеваемости был отмечен в возрастной группе детей от 3 до 6 лет — 3,4% и в возрастной группе от 0 до 2 лет — 3,8%. Самые загрязненные сульфатами атмосферные осадки отмечаются в Норильске (77 мг/л), ...выпадения серы в Норильске (9,95 т/км² в год) превысили критическое значение нагрузки в 5 раз, что является наибольшим значением среди всех загрязненных населенных пунктов Российской Федерации» [5].

Статья будет приписана к городам Норильску, Москва, Санкт Петербург и регионам Красноярский край (Норильск), Челябинская область, Забайкальский край, Курганская область, Приморский край, Нижегородская, Свердловская, Кемеровская области, Республика Бурятия, Волгоградская область, Вологодская область

Статья в журнале «Научные исследования экономического факультета. Электронный журнал». 2017 Том 9. Выпуск 1. анализирует уровень жизни граждан в регионах России — «Выделяются 5 наиболее бедных регионов, где более 25% населения имеют доходы ниже величины прожиточного минимума. Такими регионами являются Республика Тыва (42,5% бедного населения), Республика Ингушетия (31,7%), Республика Калмыкия (31,2%), Республика Алтай (26,8%), Еврейская Автономная область (25,7)» [6] Статья будет приписана к упомянутым 5 регионам.

Пример анализа демографической ситуации в муниципалитетах Центрального федерального округа — «Только семь из 281 рассмотренного муниципального образования ЦФО имели положительный естественный прирост: г. Белгород, г. Курчатова, Белгородский район, Дятьковский район, г. Железнодорожный, г. Десногорск, г.о. Обнинск». Наибольшие показатели естественной убыли населения наблюдаются в муниципальных районах Воронежской, Смоленской, Ивановской и других областях [7, с. 10]. Лидеры демографического роста (пять муниципалитетов с положительным естественным и миграционным приростом) занимают ведущие позиции в своих регионах по уровню доходов населения и наличию успешных предприятий. Белгородская область входит в пятерку регионов — лидеров в ЦФО по уровню доходов на душу населения, г. Курчатова и г. Десногорск выделяются по уровню доходов в своих регионах благодаря наличию АЭС, в г. Железнодорожном функционирует одно из крупнейших предприятий Курской области [7, с. 112].

Важный вывод статьи касается рынка труда в муниципалитетах — наибольшая доля трудоспособного населения выявлена в Ковровском районе (Владимирская обл.), Выгоничском районе (Брянская обл.), Смоленске, Курчатове (Курская обл.), Белгороде, Десногорске (Смоленская обл.), Боровском районе (Калужская обл.), Брянске, Рязани, Твери» [7, с. 109].

Интересны публикации государственных органов по регионам. Так в Бюллетене Счетной палаты в Заключении по итогам проверки ФГУП «Почты России» говорится, что «В 2016 году в сравнении с 2014 годом уменьшилась среднемесячная заработная плата отдельных категорий основного производственного персонала (операторов почтовой связи, почтальонов и сортировщиков по ряду субъектов Российской Федерации, а именно: Республике Алтай, Калининградской, Московской, Новгородской, Орловской, Тульской, Тюменской областям, Забайкальскому краю, г. Москве. Снижение составило 7,6–18,3 % по категории почтальоны; 5,7–24,4 % — по операторам связи; 8,9–15,4 % — по сортировщикам» [8].

Даже результаты первого этапа работ свидетельствуют, что обработка аналитических публикаций позволяет сформировать информационный блок, на основе которого возможно исследование ряда важных социально-экономических процессов в регионах и на местах, таких как уровень жизни граждан, рынок труда, демография и состояние здоровья детей и взрослых, экологическое состояние окружающей среды и воды, миграционные тенденции и других. Динамика процессов в значительной мере характеризует качество управления на уровне регионов и муниципалитетов.

Разработка географических компонентов для информационных систем — актуальное научное направление и прикладные исследования во всех развитых странах. Первоисточник данных в США — государственная географическая информационная система Geographic National Information System / GNIS [9].

Американский институт географических исследований разработал и поддерживает Каталог географических названий мира. Метаданные включают компонент с геологическими сведениями. Глубина и охват типов географических объектов различаются для разных стран, наиболее детально описаны США и Канада [10].

Пример разработки географического ресурса для гуманитарных наук — The Getty Thesaurus of Geographic Names. Включает детальное представление географических объектов, связанных с историей и культурой разных стран, особенно детально — США [11].

В Великобритании национальным центром географических знаний традиционно является Эдинбургский университет. Координирует исследования и поддерживает базу географических названий Научная библиотека университета. Данные по Шотландии представлены с особенной полнотой [12].

В Финляндии географическая база знаний реализована с участием Университета Тампере <http://fin.geonamebase.com/>. Один из проектов университета — сравнительный анализ криминогенной ситуации в городских и сельских районах страны. В Университете дополнили национальную базу географических названий уточнениями, характеризующими местность, по которой есть кумулятивная статистика по типам преступлений [13].

Все вышеперечисленные проекты объединяет одна специфика — на основе государственного каталога национальных географических названий университетские коллективы разрабатывают ресурсы и сервисы для поддержки географического поиска и интеграции данных и документов для системных социально-экономических исследований конкретных территорий.

ООН поддерживает портал со ссылками на национальные базы географических названий [14]. РФ в списке нет.

Проект разрабатывается как следующий этап функционального развития УИС РОССИЯ — разработка поиска по регионам РФ с использованием базы географических знаний.

Проект выполняется при поддержке Российского фонда фундаментальных исследований по гранту № 1702-12009в «База географических знаний по РФ исследований по муниципалитетам и регионам РФ».

ЛИТЕРАТУРА

1. Национальный каталог географических названий. URL: https://rosreestr.ru/wps/portal/cc_ib_data_catalog_place_names?param_infoblock_document_path=infoblock-root/cc_ib_data_catalog_place_names/index.htm (дата обращения: 20 апреля 2018 года).
2. Схема административно-территориального деления РФ. URL: <http://www.gks.ru/opendata/dataset/7708234640-okato>(дата обращения: 20 апреля 2018 года).
3. РИА Новости от 21 сентября 2017. URL : <https://ria.ru/society/20170921/1505202614.html> (дата обращения: 20 апреля 2018 года).
4. О.В. Кудрявцева, Е.Ю. Яковлева, М.С. Головин. Вопросы устойчивого развития, особенности и перспективы отечественного рынка древесного биотоплива на фоне мировых тенденций // Вестник МГУ. Серия 6. Экономика. 2016. №6. С.35.
5. Маликова О.И. Особенности изменения экологической ситуации в Российской Федерации и механизмы поддержки развития малых городов. Вестник МГУ. Серия 6. Экономика. 2017.с.с.104-106
6. Меньших Д.А. Новая концепция бедности в России как основа социальной политики. Научные исследования экономического факультета // Электронный журнал. 2017. Том 9. Выпуск 1. С.69.
7. И.Е. Калабихина, Д.Н. Мокренский. Динамика численности населения муниципальных образований центральной России // Вестник МГУ. Серия 6. Экономика. 2017. С. 10
8. Отчет о результатах контрольного мероприятия «Проверка реализации программ деятельности федеральным государственным унитарным предприятием «Почта России», а также эффективности использования бюджетных средств, выделенных предприятию в рамках государственной программы Российской Федерации «Информационное общество (2011-2020 годы)», за период 2014-2016 годов и истекший период 2017 года». Бюллетень Счетной палаты Российской Федерации. №2, 2018
9. Geographic Names Information System (GNIS). URL: <http://nhd.usgs.gov/gnis.html> (дата обращения: 10 апреля 2018 года).
10. List O - Geographic terms. URL: <http://www.americangeosciences.org/georef/list-o-geographic-terms> (дата обращения: 20 апреля 2018 года).
11. Getty Thesaurus of Geographic Names. URL: <http://www.getty.edu/research/tools/vocabularies/tgn/about.html> (дата обращения: 20 апреля 2018 года).
12. Digimap. URL: http://digimap.edina.ac.uk/webhelp/digimapsupport/about.htm#qrg/quick_ref_guides.htm (дата обращения: 20 апреля 2018 года).
13. Xingan Li. Application of Data Mining Methods in the Study of Crime Based on International Data Sources. URL: <http://tampub.uta.fi/bitstream/handle/10024/95108/978-951-44-9419-2.pdf?sequence=1> (дата обращения: 20 апреля 2018 года).
14. Searchable geographical names databases. URL: <http://unstats.un.org/unsd/geoinfo/ungegn/geonames.html> (дата обращения: 20 апреля 2018 года).

РАЗДЕЛ 2.

КУЛЬТУРА И ТЕХНОЛОГИИ. КОММУНИКАТИВНАЯ КУЛЬТУРА В ЦИФРОВУЮ ЭПОХУ