Московский государственный университет
имени М.В. Ломоносова

Научно-исследовательский вычислительный центр

Университетская информационная система

РОССИЯ

http://uisrussia.msu.ru

Октябрь 2009 года

Москва

Университетская информационная система РОССИЯ
http://uisrussia.msu.ru

Содержание

Университетская информационная система РОССИЯ (УИС РОССИЯ) с 2000 года функционирует как коллективный корпоративный ресурс - тематическая электронная библиотека и база для прикладных исследований в области экономики, управления, социологии, лингвистики, философии, филологии, международных отношений и других гуманитарных наук.

УИС РОССИЯ создана и поддерживается совместно Научно-исследовательским вычислительным центром МГУ имени М.В. Ломоносова, Экономическим факультетом МГУ имени М.В. Ломоносова и Автономной некоммерческой организацией Центр информационных исследований.

Работы по проекту ведутся с 1993 года. Исследования поддерживаются грантами Российского гуманитарного научного фонда с 1998 года, ранее поддерживались грантами Российского фонда фундаментальных исследований (1998-2004), программы «Информатизация России» Министерства науки и технологий РФ (1993 год), фонда МакАртуров, США 1993 – 1996, 2002 - 2006), фонда Форда, США (1998 - 2004), Фонда Евразия, США (2000-2004).

Доступ предоставляется бесплатно по предварительной регистрации. На конец 2009 года зарегистрированы свыше 900 коллективных пользователей (университеты, вузы, факультеты, кафедры) с правом доступа по IP-адресам и свыше 4000 пользователей - по индивидуальной регистрации из всех регионов России.

Ресурсы УИС РОССИЯ не могут использоваться в коммерческих целях. Ссылка в публикациях на первоисточник обязательна.

Справки по адресу: webmaster@mail.cir.ru

Доступ

Система размещена на сервере НИВЦ МГУ имени М.В. Ломоносова. Коллективный доступ к УИС РОССИЯ бесплатно предоставляется всем классическим университетам, вузам, институтам РАН после направления Руководителем учреждения письма-заявки на имя Директора НИВЦ МГУ Тихонравова Александра Владимировича (факс: 495 938 2136).

В письме Руководитель организации подтверждает, что ресурсы системы будут использоваться только для учебных и исследовательских целей и указывает Ответственного представителя от своей организации.

Ответственный представитель организации сообщает IP-адреса классов коллективного доступа и электронные адреса преподавателей и сотрудников для индивидуального доступа.

Пароль для индивидуального доступа предоставляется после заполнения регистрационной формы.

Информационные ресурсы УИС РОССИЯ

Общий объем ресурса – около 40 Гбайт, около 3 млн. документов и более 400 000 статистических таблиц из 100+ источников – информационных партнеров проекта.

Интегрированная коллекция

УИС РОССИЯ формируется из электронных версий первоисточников, получаемых по Соглашениям о сотрудничестве между Научно-исследовательским вычислительным центром МГУ имени М.В. Ломоносова и правообладателями ресурсов. Поддерживаются следующие представленные в ретроспективе и обновляемые на регулярной основе коллекции:

- нормативные документы федерального уровня – законы, указы, распоряжения, постановления;
- стенограммы пленарных заседаний и постановления Государственной Думы Федерального Собрания РФ;
- международные договоры РФ;
- сборники и аналитические доклады Федеральной службы государственной статистики (Росстат);
- доклады / мониторинги государственных органов - Министерства экономического развития, Центрального банка, Счетной палаты, Совета Федерации и других;
- аналитические доклады, публикации и статистические ресурсы российских и международных исследовательских центров – Ассоциации независимых центров экономического анализа, Экономической экспертной группы, Центра фискальной политики, Независимого института социальной политики, Всемирного банка, Леонтьевского центра и других;
- научные издания – «Социологический журнал», «Демоскоп», «Квантиль» и другие;
- издания СМИ – газеты, «Ведомости», «Известия», «Коммерсантъ», «Независимая газета», «Новая газета», «Поиск»;
- журнал «Эксперт»;
- зарубежные источники – базы данных, доклады, архивы международных организаций и центров по изучению России, публикации по экономике и социологии;
- архив выборной статистики Центральной избирательной комиссии.

Тематические разделы УИС РОССИЯ

По наиболее востребованным направлениям исследований и тематикам учебных программ разработаны предметно-ориентированные разделы, реализуются специальные сервисы, нацеленные на формирование информационно-аналитической среды «рабочее место специалиста» *(professional workbench)*.

Российская Федерация: социально-экономическая статистика

Наряду с базами данных, сформированными на основе публикаций Росстата, в системе хранятся и поддерживаются электронные версии публикаций Росстата - более 200 изданий (27 названий ежегодных статистических сборников, отраслевые и другие сборники). Ретроспектива изданий – с 1996 года.

Методологическое сопровождение таблиц организовано как многоуровневый электронный справочник с развитыми гипертекстовыми связями. Краткие пояснения составлены с использованием материалов соответствующих статистических сборников и привязаны к конкретным таблицам. Развернутые методологические положения детально описывают стандарты российского статистического учета. В гипертекстовом режиме доступен Глоссарий статистических терминов.

По всем публикациям Росстата возможен поиск с использованием терминов Тезауруса УИС РОССИЯ, рубрикаторов, названия публикации.

По коллекции сборников Росстата доступна навигация с использованием градации таблиц, принятой в полиграфической версии.

Российская Федерация: население и уровень жизни

Тематический раздел «Российская Федерация: население и уровень жизни» интегрирует данные из нескольких источников, характеризующие аспекты социально-демографического развития России – численность населения, естественное и механическое движение, уровень жизни и развитие социальной сферы, правонарушения и социальное неблагополучие. Источник данных – ежегодные издания Росстата, а также база "Паспорта городов Российской Федерации" с данными по 1050 городам.

Ресурс включает более 100000 аналитических таблиц в формате HTML. Общий объем базы – 2,4 Гбайт. Для удобства последующей обработки данных все таблицы доступны в формате MS Excel и zip-архиве.

Доступен толковый словарь демографических терминов, описывающий содержание более 570 понятий. Словарь создан на базе первоисточника: Демографический понятийный словарь / Под ред. д.э.н. проф. Рыбаковского Л.Л. - М.: Центр социального прогнозирования, 2003.

В разделе реализована методика интеграции статистических показателей федерального и регионального уровней из разных источников на основе классификатора. Методика разработана коллективом проекта. Классификатор включает 22 раздела, в том числе: общие показатели воспроизводства населения; браки и разводы; правонарушения; состояние здоровья населения; социальное обеспечение и социальная помощь; образование; условия труда и ряд других разделов. Классификатор обеспечивает единую систему навигации по всем таблицам федеральной и региональной статистики, включенным в базу, независимо от первоисточника.

Ретроспектива данных с 1996 года; навигация по таблицам с данными за каждый год может осуществляться двумя способами, связанными между собой: через сводное оглавление с разделами и оглавление с источниками данных.

Для систематизированного представления данных муниципального уровня используются следующие решения:

- классификатор показателей,
- алфавитный список муниципальных образований,
- иерархические меню «федеральный округ – регион – муниципалитет».

Эта часть ресурса представлена двумя типами таблиц: таблицей значения показателя по нескольким городам, расположенным на территории одного субъекта федерации; и сводными таблицами по каждому городу, содержащими все доступные показатели из одного раздела классификатора. Ретроспектива данных – с 1970 года.

Российская Федерация: аграрно-промышленный комплекс

Ресурс предназначен для проведения комплексных исследований развития аграрного сектора в России и включает данные о производстве сельскохозяйственной продукции, материально-технической базе сельского хозяйства, сельскохозяйственных организациях, ценообразовании на продукцию сельского хозяйства, трудовом потенциале аграрного сектора, а также основных показателях социального развития села.

Источники данных – публикации Росстата и издания региональных статистических органов, в том числе:

- база данных "Паспорта городов Российской Федерации" ГМЦ Росстата;

- данные в разрезе страны за 1996–2005 годы из 96 сборников Росстата;
- данные в разрезе регионов за 1996–2005 годы из 76 сборников Росстата;
- данные в разрезе административных районов из 178 статистических сборников, изданных в регионах. *Этот уникальный информационный фонд предоставлен авторами проекта «A Troubled Realm: Russian Agriculture's Spatial Constraints, Variance, and Prospects for Revival», проведенного сотрудниками Института географии РАН совместно с Рэдфордским университетом, США, и Калифорнийским университетом в Сан Диего, США, по гранту Национального научного фонда США в 2000 -2002 годах.*

База включает более 19000 аналитических таблиц, 55% таблиц содержат данные в разрезе муниципальных образований (городов и административных районов), 20% таблиц – в разрезе субъектов Российской Федерации, 25% – данные федерального уровня. Общий объем ресурса составляет 1,6 Гбайт. Все таблицы доступны и в формате MS Excel.

Таблицы базы данных имеют методологическое сопровождение, а также ссылки на тематический Глоссарий. К каждой таблице указаны полные реквизиты первоисточника, а также ссылка на сайт соответствующего ведомства.

В разделе реализована методика интегрированного представления показателей из нескольких статистических первоисточников. Методика разработана коллективом проекта.

Данные систематизированы на основе единого классификатора показателей, в основу которого положен действующий рубрикатор Федеральной программы статистических работ Росстата. Классификатор включает 27

разделов, в числе которых: основные показатели сельского хозяйства; земельные ресурсы; предприятия и организации; животноводство; растениеводство; инвестиции, основные фонды и ввод в действие производственных мощностей; цены и тарифы; внешнеэкономическая деятельность и др. Внедрение классификатора позволило создать единую систему навигации по всем первоисточникам федеральной и региональной статистики, включенным в базу данных. Помимо этого имеются средства навигации с привязкой к первоисточникам (по названиям статистических сборников).

Для систематизации данных муниципального уровня использованы:

- классификатор показателей,
- алфавитный список муниципальных образований,
- иерархические меню «федеральный округ – регион – муниципалитет».

Ввиду большого объема муниципальных данных, использованы различные формы их представления. Для целей сравнительного анализа можно получить таблицу значений показателя по нескольким муниципалитетам, расположенным на территории одного субъекта федерации. По каждому городу и административному району имеются также сводные таблицы, содержащие все доступные показатели, относящиеся к какому-либо разделу классификатора. Переход от одного типа таблиц к другому достигается с помощью системы гиперссылок.

Ресурс содержит данные по 2005 год, обновление временно не производится. Ищем партнеров.

Права человека: документы международных организаций

Цель раздела - поддержание многофункциональной информационной системы для учебных программ, исследований, просвещения граждан в области прав человека и социальной защиты.

Ресурс поддерживается на общедоступном сайте http://www.echr-base.ru

Модуль по правам человека включает:

- полный архив документов Европейского Суда по правам человека (на английском и французском языках),
- коллекцию документов Суда, доступных в переводе на русский язык,
- документы Совета Европы, ООН, СНГ по правам человека на русском и английском языках,
- развитый справочный блок на русском языке по статусу и процедуре деятельности Европейского Суда по правам человека,
- применимое национальное законодательство – кодексы, нормативно-правовые акты РФ, рассматривавшиеся по жалобам против РФ;
- книги и публикации исследовательских центров,
- библиографию изданий по правам человека.

Источники документов:

- официальные публикации ЕСПЧ, полученные из первоисточника - Отдела публикаций Секретариата Европейского Суда в Страсбурге (HUDOC DVD), обновление – 3 раза в год;

- переводы на русский язык, получены из официальных изданий и от переводчиков;
- учебные материалы юридических факультетов университетов и вузов, документы исследовательских центров и правозащитных организаций, научные статьи и материалы СМИ, получены по Соглашениям о сотрудничестве с правообладателями ресурсов.

Предполагается поэтапно интегрировать в систему все основные публикации по вопросам защиты прав человека, издаваемые в России.

Архив документов ЕСПЧ (HUDOC CD-ROM), в версии на ноябрь 2009 года включает более 40000 документов на английском и французском языках. Архив загружен в информационную систему с интерфейсом на русском языке и развитым поисковым механизмом — поиск по статьям Европейской Конвенции, по фамилии заявителя, по государству-ответчику, по типу документов, по ключевым словам и др. Каждый документ сопровождается карточкой на русском языке.

В отдельный раздел вынесены документы Суда по жалобам против России. Раздел включает специальный сервис - доступ:

- к прецедентам Суда, упоминаемым в деле;
- статьям Европейской Конвенции и другим документам Совета Европы;
- применимому национальному законодательству - соответствующим разделам Конституции и нормативно-правовых актов РФ;
- переводам документов Суда на русский язык и комментариям.

Справочный раздел сайта о деятельности Европейского Суда - поддерживает ссылки на отчеты и материалы, представленные на официальном сайте Суда. Раздел "Публикации" содержит полные тексты изданий партнеров проекта. Поддерживается Библиография изданий по правам человека, опубликованным на русском языке.

Модуль по Европейской социальной хартии содержит документы о ратификации Хартии Российской Федерацией, справочную информацию по Хартии и Европейскому Комитету по социальным правам – статус, регламент, процедуры контроля и коллективных жалоб, а также тексты национальных докладов и заключений Комитета по национальным докладам по странам – членам Хартии.

Классика Российского права

Раздел содержит копии полных текстов документов, монографий, учебников, статей по вопросам права, изданным в XIX – начале XX века, в том числе все 16 томов "Свода законов Российской империи" (издание 1912 года), «Судебные уставы» 1864 года. Документы предоставлены компаниями «Консультант» и «Гарант».

Базы данных и on-line анализ

В отдельный раздел УИС РОССИЯ выделены базы данных и сервисы для анализа в он-лайн режиме.

База «Регионы России: интегрированная база по бюджетной и социально-экономической статистике»

Блок социально-экономических показателей в разрезе субъектов РФ включает данные ежегодных сборников Росстата «Регионы России», «Экономическая активность населения России (по результатам выборочных обследований)», «Цены в России», «Демографический ежегодник России», «Семья в России», «Российский статистический ежегодник», «Торговля в России», «Труд и занятость в России», «Здравоохранение в России», «Жилищное хозяйство и бытовое обслуживание населения в России». Представлено порядка 2000 показателей, большинство которых - в ретроспективе с 1990 года, некоторые – с 1970 года. Показатели структурированы в соответствии с рубрикацией, используемой в сборниках «Регионы России» с 2005 года. Показатели из сборников за разные годы, начиная с 2000 года, сведены во временные ряды. Показатели снабжены методологическими пояснениями Росстата и «привязаны» к формам статистической отчетности, дополненным инструкциями по их заполнению.

Блок показателей бюджетной статистики формируется на основе данных ведомственной статистики, размещенных на официальных сайтах Минфина и Росказны. Данные представлены в ретроспективе с 1998 года и включают как плановые показатели, так и фактические данные. Для бюджетных данных с 2006 года реализована процедура агрегирования показателей по группам, статьям и подстатьям

экономической классификации, видам и целевым статьям расходов. Интерфейс сводной реляционной базы обеспечивает пошаговый выбор:

- Региона,
- Года/периода,
- Показателей бюджетной и социально-экономической статистики из соответствующих рубрик,
- Вида бюджета.

База «Города России»

База реализована на основе публикации ГМЦ Росстата "Паспорта городов", содержит значения 699 основных показателей социально-экономического развития по 1051 городу, имеющему статус муниципального образования. Ретроспектива показателей – 1991 год, некоторые показатели доступны с 1970 года. Интерфейс реляционной базы предполагает пошаговый выбор:

- Города,
- Года/периода,
- Показателей социально-экономической статистики.

База «Муниципальные образования»

База разработана на основе публикации Росстата «Муниципальная статистика. Паспорт муниципального образования», содержит данные по 25020 муниципальным образованиям по 349 показателям за 2005-2008 годы.

On-line анализ

В базах реализован комплекс инструментов для анализа и прогноза показателей социально-экономического развития, включая процедуры вычисления:

- вторичных переменных вариационных рядов и рядов динамики;
- уравнения регрессии и стандартных показателей корреляции;
- прогнозных значений показателя на ближайшие пять лет с использованием взвешенной регрессии.

Среди других функциональных возможностей – отбор по критерию, построение графиков, диаграмм и картограмм. Все вычисляемые вторичные переменные снабжены справкой и ссылками на соответствующие публикации в составе УИС РОССИЯ.

Реализован интерактивный обучающий Практикум по использованию процедур анализа и прогноза.

Обновление баз производится ежегодно по мере публикации данных Росстатом и Федеральным казначейством

База «Оперативная статистика»

База содержит данные, публикуемые в региональном приложении к ежемесячному сборнику «Социально-экономическое положение России» и в публикации Росстата «Информация для ведения мониторинга социально-экономического положения субъектов Российской Федерации», включает 1140 показателей в ретроспективе с января 2006 года.

Данные обновляются ежемесячно.

Сервис всех баз данных включает средства поддержки работы с таблицами, прежде всего конструктор таблиц. Конструктор позволяет формировать и форматировать таблицу, в том числе задавать содержание строк и столбцов, выбирать вспомогательные коды. Таблицы могут быть экспортированы в формат MS Excel, представлены в виде .csv файлов.

База «НОБУС»
(Национальное обследование благосостояния и участия населения в социальных программах)

База создана на основе данных обследования, проведенного Росстатом при техническом содействии Всемирного Банка в 2003 году.

Обследование включало 227 вопросов, сгруппированных в 13 разделов: общая информация; описание домохозяйства; семейное положение и образование; занятость; виды назначенных пенсий; виды назначенных пособий; категории льготополучателей и виды льгот; здоровье и медицинское обслуживание; социальная помощь; жилищные условия; подсобное хозяйство; расходы; доходы. В ходе обследования собраны данные о потреблении и доходах домохозяйств, а также об их демографических характеристиках и занятости, доступности здравоохранения, образования и социальных программ, субъективных оценках собственного благосостояния. Реализованы сервисы:

- доступ к данным через иерархическое дерево показателей, состоящее из 13 верхних уровней (разделов обследования), 227 вопросов;
- построение частотных таблиц, отражающих количество ответивших определенным образом на вопрос, процент

от числа опрошенных, процент от значимых ответов на вопрос, накопленный процент. Пользователь может загружать частотные таблицы в формате MS Excel;

- процедура критериального отбора данных. Пользователь может фильтровать исходные данные по определенному критерию (т.е. выбирать определенные варианты ответа на вопрос) при загрузке и построении частотных таблиц и сохранении данных;
- расчет некоторых статистических показателей: средняя арифметическая; мода; медиана; дисперсия; коэффициент асимметрии; эксцесс; размах; минимум; максимум и др.;
- загрузка части показателей с использованием критериального отбора на пользовательский компьютер из исходной базы в формате .cvs, распознаваемом MS Excel;
- загрузка полной версии в исходном формате, читаемом статистическими пакетами SPSS, STATA;
- получение краткого описания методологии обследования в формате pdf.

Данные НОБУС предоставлены проекту УИС РОССИЯ Московским представительством Всемирного банка.

Академический сервис

Разработка решений, нацеленных на повышение функциональности системы и обеспечение услуг, ориентированных на профессиональные потребности пользователей, – академический сервис - специальное направление проекта УИС РОССИЯ.

Реализованные технологические процедуры и лингвистические технологии обеспечивают техническую и библиографическую обработку и анализ содержания электронных документов и данных на входе в систему. Процедуры и технологии адаптированы для обработки документов каждой из коллекций УИС РОССИЯ, чем обеспечивается содержательная интеграция коллекций и возможность сквозного поиска по всем источникам.

Включение каждой новой коллекции предполагает настройку процедур на соответствующий формат представления документов и обновление лингвистических средств.

Выполнение описанного трудоемкого комплекса работ обеспечивает дополнительную функциональность УИС РОССИЯ и на этапе поиска, и на этапе исследования. При поиске пользователь экономит время, благодаря более рациональной процедуре отбора, обработки и организации документов и данных, возможности оперативно уточнить запрос.

Пользователям УИС РОССИЯ доступна процедура обновления типовых запросов в автоматическом режиме на личной странице пользователя, включая информирование о новых поступлениях по электронной почте.

Поисковые инструменты

В разделе «Интегрированная коллекция» в текущей версии свыше 3 млн. документов и более 400 000 статистических таблиц из 100+ источников. В разделе доступны как стандартные поисковые инструменты, так и специальный сервис для уточнения запроса на основе содержательного анализа результатов выдачи. Эта процедура позволяет оперативно и интерактивно уточнять запрос, сокращая время на поиск нужных документов.

Для поиска по Интегрированной коллекции доступны два рубрикатора:

- Рубрикатор 1 разработан для УИС РОССИЯ и предназначен для обработки и поиска, прежде всего, нормативно-правовых документов. Содержит 180 рубрик, три уровня вложенности;
- Рубрикатор 2 представляет собой Верхний уровень Тезауруса LIV/Legislative Indexing Vocabulary, разработан Исследовательской службой Библиотеки Конгресса США. Содержит 80 рубрик.

Для отдельных коллекций дополнительно доступны специальные классификаторы. Так, для коллекции статей научных изданий и публикаций исследовательских центров дополнительно доступен поиск по рубрикатору JEL/Journal of Economic Literature Classification System.

Наиболее гибкий способ поиска и уточнения запроса - навигация по связям Общественно-политического тезауруса (далее - Тезаурус УИС РОССИЯ). Тезаурус – словарь понятий с указанием отношений и связей между понятиями (иерархических («выше–ниже»), ассоциативных и синонимичных). В Общественно-политическом тезаурусе более

33 000 понятий, свыше 80 000 синонимов, около 1 000 000 связей. Переход по связям Тезауруса обеспечивает возможность оперативного и интерактивного уточнения запроса на всех этапах поиска.

Наиболее полезная для специалиста функциональная особенность, реализованная в «Интегрированной коллекции», - получение по запросу тематической подборки из статистических таблиц сборников Росстата, Центрального банка, Министерства экономического развития, Счетной палаты и аналитических документов (доклады, обзоры, научные статьи и другие публикации). Например, таблиц с данными, характеризующими уровень жизни населения, и документов на эту тему, в том числе публикаций с анализом данных Росстата, характеризующим уровень жизни.

Информеры

В системе реализован комплекс т.н. информеров – таблиц, в которых представлены результаты запроса, сгруппированные по разным поисковым признакам – тематика, дата, регион. Информер по тематике представлен в виде таблицы со списком понятий Тезауруса, наиболее характерных для массива документов, полученных в результате исполнения запроса. Список терминов упорядочен по убыванию значимости.

Информер по дате представляет результаты запроса, сгруппированные по годам/месяцам издания документов.

Информер по регионам, соответственно, представляет результаты, сгруппированные по субъектам РФ.

Для коллекции «Стенограммы пленарных заседаний Государственной Думы ФС РФ» реализованы дополнительные информеры – дата заседания, фамилия депутата.

Технология подготовки статистических данных

Для коллекций статистических данных производится дополнительный комплекс исследований и трудоемкие научно-вспомогательные и технические работы, в том числе:

- перевод таблиц в формат MS Excel с одновременной проверкой правильности оформления таблиц;
- сопровождение таблиц соответствующими методологическими пояснениями;
- перевод статистики, публикуемой Росстатом, в формат баз данных с предварительной проверкой на сопоставимость данных и необходимым справочным сопровождением;
- процедуры для визуализации данных с помощью графиков и картограмм.

Работы выполняются в автоматизированном режиме под контролем оператора.

Наукоемкая часть работ – формирование баз данных на основе публикаций Федеральной службы государственной статистики и Федерального казначейства. Наиболее востребованные сборники Росстата - «Регионы России» и «Социально-экономическое положение России» составили основу базы «Регионы России». В базе показатели представлены в виде ежегодно обновляемых временных рядов в ретроспективе с 1990 года (по ряду показателей).

Данные бюджетной статистики федерального и регионального уровня также переводятся в формат реляционной базы. Процесс работы включает анализ изменений в бюджетной классификации за последние годы, поддержание в актуальном состоянии классификаторов и модификация базы с учетом изменений. Одновременно отслеживаются изменения в

системе административно-территориального деления страны, корректируются географические названия.

В результате сформирован и поддерживается комплекс статистических ресурсов для поддержки исследований по широкому кругу социально-экономических проблем с использованием методов системного и сравнительного анализа на уровне страны, регионов, муниципалитетов. В числе :

- база, в которой интегрированы социально-экономические данные Росстата по регионам и бюджетные данные Федерального казначейства по регионам;
- база социально-экономических показателей по городам ;
- база на основе социально-экономической статистики по 25020 муниципалитетам;
- база по оперативной социально-экономической статистике по регионам с ежемесячным обновлением.

Базы обеспечивают поиск и анализ каждого показателя (или группы показателей) по состоянию на определенную дату/период времени и с конкретной территориальной привязкой. Доступны средства визуализации данных с помощью графиков, диаграмм, картограмм.

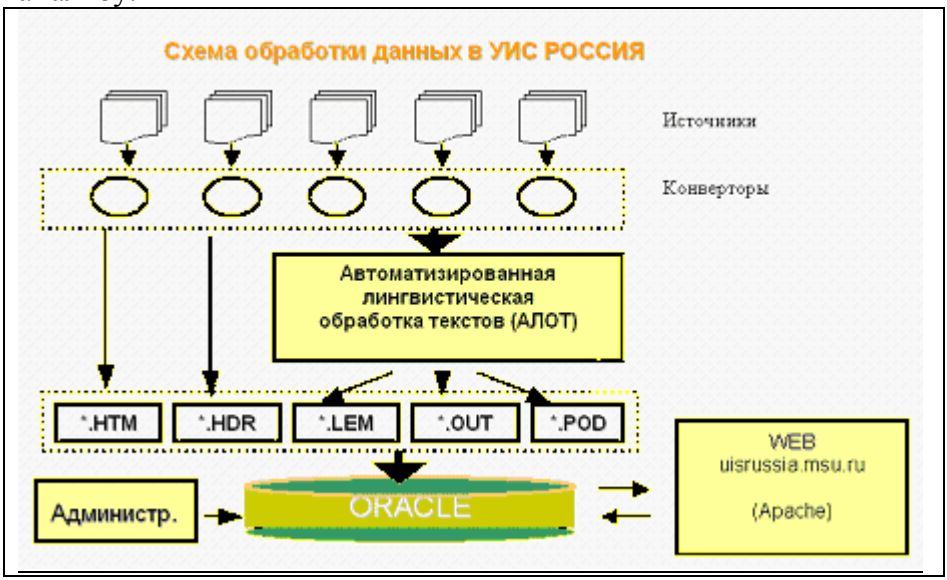
В базах данных реализован комплекс инструментов для анализа и прогноза показателей социально-экономического развития, включая процедуры для вычисления:

- вторичных переменных вариационных рядов и рядов динамики;
- уравнения регрессии и стандартных показателей корреляции;
- прогнозных значений показателя на ближайшие пять лет с использованием взвешенной регрессии.

Все вычисляемые вторичные переменные снабжены справкой (определение, алгоритм (формула) вычисления, области, особенности и примеры применения) и ссылками на соответствующие публикации в составе УИС РОССИЯ. Разработан интерактивный обучающий Практикум по использованию процедур анализа и прогноза.

Общая технология обработки документов

УИС РОССИЯ реализована как интегрированный ресурс на основе технологии автоматизированной лингвистической обработки текстов (АЛОТ). Технология разработана в рамках проекта и стала результатом комплекса фундаментальных и прикладных исследований по лингвистике и информационному анализу.



На первом этапе обработки документов на входе в систему несколько программ последовательно выполняют следующие процедуры:

- перевод данных, поступающих из разных источников, в единообразный формат хранения;
- библиографическая обработка источников (краткая форма), формирование файла метаданных;
- библиографическая обработка документов и статистических таблиц, приписывание библиографического описания источника к каждому документу/таблице.

Для сложных по структуре (содержат таблицы, графики, рисунки, формулы, сноски, примечания и т.д.) и поэтому трудоемких для обработки документов - докладов/бюллетеней/вестников государственных ведомств и публикаций исследовательских центров - разработан специальный конвертор. Конвертор включает несколько модулей и содержит макросы, использующие алгоритм кластеризации текста на основе детального анализа его структуры. Один из макросов настроен на автоматизированную разметку документов - в тексте специальным образом выделяются заголовки первых трех уровней, таблицы, сноски и т.д. Конвертор обрабатывает как документы Microsoft Word, так и HTML и создает директорию с htm-файлами определенного формата, файлами метаданных, таблицами в формате Microsoft Excel. Обработка каждого нового источника требует настройки конвертера на формат источника. Применение конвертора значительно уменьшает объем ручной работы оператора.

Следующий этап обработки - содержательный анализ документов и статистических таблиц на базе нескольких лингвистических процессоров (технология АЛОТ). В автоматическом режиме производится:

- систематизация/классификация документов и заголовков статистических таблиц по рубрикаторам;
- терминологический анализ и индексирование документов и заголовков статистических таблиц и названий показателей по Тезаурусу УИС РОССИЯ;
- рубрицирование статей научных изданий и публикаций исследовательских центров дополнительно по рубрикатору JEL/Journal of Economic Literature Classification System (считается международным стандартом для обработки научных публикаций по экономике и социологии);
- аннотирование полнотекстовых документов в виде фрагментов текста, раскрывающих основные темы документа.

Технология АЛОТ адаптирована для обработки всех основных типов документов жанра «деловая проза» на русском языке и для обработки научных публикаций и материалов СМИ на английском языке. Результаты обработки – метаданные - загружаются в информационную систему на СУБД Оракл и используются для поддержки развитого механизма сквозного поиска по Интегрированной коллекции.

Перспективы развития УИС РОССИЯ. Распределенная сеть ресурсов по России

Содержательное развитие УИС РОССИЯ ведется с учетом пожеланий целевой аудитории – исследователей и преподавателей гуманитарных факультетов университетов РФ.

Результаты опросов пользователей говорят о востребованности публикаций и других ресурсов исследовательских организаций и научных центров. В последние годы большинство российских и международных научных центров рассматривают создание и поддержание Интернет-ресурсов по своей тематике как одно из основных направлений деятельности.

Коллекции научных организаций относятся к наиболее качественным источникам с точки зрения содержания, ретроспективы, полноты, регулярности обновления. Число таких ресурсов растет, и коллектив УИС РОССИЯ предложил партнерам интегрировать коллекции в рамках распределенной сети.

В состав распределенной сети уже включены журнал «Демоскоп» и другие издания ГУ-ВШЭ, журнал «Квантиль» Российской экономической школы, публикации «Леонтьевского центра», доклады «Всемирного банка», газета «Коммерсант».

Предполагается поэтапно интегрировать в составе распределенной сети основные качественные источники по тематике УИС РОССИЯ. Возможны разные схемы участия и формы взаимодействия по поддержанию распределенной сети.

Наиболее востребованные ресурсы в составе УИС РОССИЯ — разделы, реализованные на основе государственной статистики РФ. Поддержание и развитие этих ресурсов, реализация дополнительных аналитических сервисов - основное направлением работы коллектива. Разработана первая версия онтологии «государственное и муниципальное управление» для интеграция статистических данных из разных источников для комплексного анализа.

Коллектив УИС РОССИЯ разрабатывает специальный практикум по прикладному анализу для задач учебного процесса в соответствие с новыми федеральными государственными образовательными стандартами высшего профессионального образования. Стандарты по направлению «Экономика» в части профессиональной деятельности и бакалавров, и магистров предусматривают умение работать с данными и способность использовать количественные и качественные методы для проведения научных исследований и управления бизнес-процессами, владение методами экономического анализа поведения экономических агентов и рынков в глобальной среде и методами стратегического анализа. Практикум ориентирован на овладение профессиональными компетентностями и навыками использования потенциала государственной статистики для конкретной задачи. Созданная в УИС РОССИЯ современная статистическая база позволит приблизить изучаемые в вузах теоретические аспекты теории вероятностей и социально-экономической статистики к практике, дополнить прикладными курсами, включая решение экономико-математических примеров и задач по анализу реальных данных.

Из опыта зарубежных стран. Справка

Поддержание сети тематических информационных центров на базе университетов — практика научных сообществ всех развитых странах мира.

В последние годы ресурсы университетов все более востребованы органами власти, бизнесом, обществом и постепенно становятся составной частью национальной информационной инфраструктуры нового поколения.

Компьютерные технологии увеличивают возможности исследователей в гуманитарных науках - расширяют круг источников и обеспечивают оперативный доступ к ним, машиночитаемая форма документов и данных поддерживает приемы обработки и анализа, недоступные при работе с печатными изданиями, в том числе использование математических методов анализа.

Вместе с тем, большой и постоянно возрастающий информационный поток ставит задачу отбора, систематизации и организации документов и данных, разработки информационных систем и сервисов для поддержки конкретных исследований. Потенциал и результаты научных проектов в значительной мере определяются качеством электронных ресурсов, поэтому формирование базы электронных ресурсов для полноценных исследований – важная задача каждого научного коллектива. Создание и поддержание качественного информационного ресурса – долгосрочная, трудоемкая и дорогостоящая деятельность.

Практика университетских сообществ мира доказывает рациональность коллективной инфраструктуры в виде корпоративной сети информационных центров, деятельность

которых направлена на целенаправленное и скоординированное формирование информационной базы для исследований и образовательных программ.

В большинстве стран на базе крупного университета или института создан коллективный информационный орган, который играет роль национального координатора. В задачи центра входят три основных направления.

Первое - изучение информационных потребностей специалистов - представителей всех направлений гуманитарных наук и определение информационных приоритетов с учетом интересов большинства пользователей, маркетинг, переговоры с правообладателями ресурсов, юридическое оформление и получение архивов, обеспечение полноты и целостности коллекций.

Второе направление – формирование и поддержание электронной библиотеки, разработка и реализация решений, направленных на повышение функциональности ресурсов. Речь идет о прикладных исследованиях и разработке программных процедур по предварительной технической и содержательной обработке электронных документов и данных, переводу в форматы, удобные для анализа. Цель этого комплекса работ - реализация дополнительных услуг, ориентированных на профессиональные потребности пользователей, – академический сервис.

Третье направление – образовательное - разработка учебных программ и проведение обучения по новым для гуманитарных наук методам исследований – количественному анализу данных с применением математических методов.

В последние годы усилиями научного сообщества в США и странах Западной Европы целенаправленно формируются национальные сети ресурсов для продвижения

статистического образования. Работы ведутся в рамках национальных программ создания информационной инфраструктуры 21 века. Такие программы приняты во всех развитых странах мира. Работы ведутся совместными усилиями научного сообщества и правительства при активной и координирующей роли национальных научных фондов.

На этапе построения «общества знаний» ресурсы университетского сообщества становятся составной частью национальной информационной инфраструктуры, а сами университеты стали ведущей движущей силой информационного развития общества и продвижения новых технологий и практик управления страной.

Подписано в печать 21 октября 2009 года. Формат 60x84/16. Бумага офс. №1. Печать ризо. Усл.печ.л.2. Тираж 70 экз. Заказ №

Участок оперативной печати НИВЦ МГУ.
119992 Москва, Воробьевы горы,
НИВЦ МГУ имени М.В. Ломоносова