

ISBN 978-5-19601-104-3

Федеральное государственное бюджетное учреждение науки
**ИНСТИТУТ КОСМИЧЕСКИХ ИССЛЕДОВАНИЙ
РОССИЙСКОЙ АКАДЕМИИ НАУК**

Федеральное государственное бюджетное учреждение науки
**ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР им. А.А. ДОРОДНИЦЫНА
РОССИЙСКОЙ АКАДЕМИИ НАУК**

**ИНФРАСТРУКТУРА
НАУЧНЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ
И СИСТЕМ**

Сборник избранных научных статей

Под редакцией

доктора техн. наук **Е.Б. КУДАШЕВА**,
доктора физ.-матем. наук **В.А. СЕРЕБРЯКОВА**

Том II



Москва
2014

УДК [002:004.9] (063)
ББК [73+32.973.233]я43

Инфраструктура научных информационных ресурсов и систем. Сборник избранных научных статей. Труды Четвертого Всероссийского симпозиума (С.-Петербург, 6–8 октября 2014 г.). Под ред. Е.В. Кудашева, В.А. Серебрякова. В 2-х тт. Т. 2. М.: ВЦ РАН, 2014.

Симпозиум проводится ежегодно с 2011 г. по Плану научных конференций Отделения математических наук РАН. В 2011 г. и 2012 г. Симпозиум проводился в С.-Петербурге при поддержке РФФИ, в 2013 г. – в Абхазии, г. Сухум – при поддержке Академии наук Абхазии. В 2014 г. Симпозиум проводился в С.-Петербурге при поддержке РФФИ на базе Петербургского Отделения Математического института им. В.А. Стеклова РАН – ПОМИ РАН.

Научная программа Симпозиума «**Инфраструктура научных информационных ресурсов и систем**» ориентирована на рассмотрение проблем и перспектив развития информационно-телекоммуникационных систем; методов, технологий и средств применительно к доступу, хранению и интеллектуальному анализу данных в различных областях фундаментальной науки, разработки информационных систем для научных исследований.

Основные цели Четвертого Симпозиума: методы и технологии интеграции электронных коллекций; взаимодействия информационных ресурсов и формирования электронного документного пространства научных исследований и инноваций, развитие электронных библиотек.

The symposium is held annually since 2011 on the Plan of scientific conferences Department of Mathematical Sciences of RAS. In 2011 and 2012 Symposium was held in St. Petersburg and was supported by RFBR, in 2013 – in Abkhazia, Sukhum – with the support of the Academy of Sciences of Abkhazia. In 2014, the Symposium is held in St. Petersburg on the basis of the St. Petersburg Branch of the Mathematical Institute of the Academy of Sciences – PDMI RAS and is supported by RFBR.

The scientific program of the Symposium is oriented to the infrastructure of scientific information resources and systems geared to the problems and prospects of development of information and telecommunication systems; methods, tools and technology with respect to access, storage, and data mining in various fields of basic science, development of information systems for research.

The main objectives of the Fourth Symposium: methods and techniques of integration of digital collections; interaction of information resources and the generation of the electronic document space research and innovation, the development of digital libraries.

Рецензенты: Г.Н. Заварза, К.Б. Теймуразов

Научное издание

© Федеральное государственное бюджетное учреждение науки
Вычислительный центр им. А.А. Дородницына Российской академии наук, 2014

WPS-СЕРВИСЫ ПРОСТРАНСТВЕННОГО АНАЛИЗА СОСТОЯНИЯ ОКРУЖАЮЩЕЙ СРЕДЫ И ПРИРОДНЫХ РЕСУРСОВ

Р.К. Фёдоров, А.С. Шумилов

Институт динамики систем и теории управления СО РАН,
Иркутск, Россия
fedorov@icc.ru

В статье описывается разработка среды распределенных вычислений на основе стандарта Web Processing Service, который унифицирует использование Web-сервисов, предоставляющих услуги пространственной обработки растровых и векторных данных.

Ключевые слова: OGC, WPS, JavaScript, V8, Calipso

The article describes the development of the WPS-based (Web Processing Service) distributed process environment, which simplifies the use of different web-services, offering processing and analysis of the geospatial data.

Keywords: OGC, WPS, JavaScript, V8, Calipso

7. Введение

Решение задач анализа состояния окружающей среды и природных ресурсов, как правило, требует применения нескольких программных систем (ПС), разрабатываемых специалистами различных предметных областей. Например, моделирование загрязнения атмосферного воздуха населенных пунктов используются методы инвентаризации источников загрязнений (печное отопление, автотранспорт и т.д.) и моделирования распространения загрязнений. Эти методы являются достаточно сложными и часто реализовываются разными коллективами. Взаимодействие коллективов осуществляется редко в частности из-за сложности использования и интеграции пакетов программ. В итоге при решении сложных задач решается одна или несколько подзадач на

должном уровне в зависимости от ресурсов коллектива, а остальные подзадачи решаются упрощенным способом.

На сегодняшний день активно развивается взаимодействие между программными системами (интероперабельность) через Интернет, используя стандарты Open Geospatial Consortium (OGC) [1]. Одним из перспективных стандартов OGC является Web Processing Service (WPS) [2], который унифицирует использование сервисов обработки пространственных данных (ПД) через Интернет. Например, это могут быть сервисы, реализующие пространственную обработку растровых и векторных данных, геомоделирование, методы статистики. Достоинствами данного стандарта являются простота, возможность предоставления метаданных, поддержка длительного выполнения сервисов и т.д. В области обработки ПД он позволяет объединить и интегрировать программные системы, созданные разными разработчиками, для решения сложных задач, требующих объединения специалистов в различных предметных областях. В статье рассматривается подход, позволяющий объединить распределенные данные и WPS-сервисы для решения задач.

8. Распределенная вычислительная среда

Для эффективного взаимодействия WPS-сервисов и данных требуется организация распределенной вычислительной среды. Вычислительная среда должна предоставлять возможность хранения данных и выполнения алгоритмов (программ) над этими данными. Соответственно распределенная вычислительная среда должна пользователю и WPS-сервисам предоставлять доступ и возможность размещения в Интернет исходных данных, промежуточных данных и результатов вычислений (обработки) и выполнение распределенных WPS-сервисов.

В ИДСТУ СО РАН разрабатывается геопортал, который реализует основные функции среды. Геопортал предоставляет пользователям функции хранения данных в следующем виде:

1) файлов в представляемой пользователю директории файловой системы геопортала на базе системы хранения данных (СХД) SAN ReadyStorage 3994;

2) реляционных таблиц в СУБД PostgreSQL.

Геопортал предоставляет файловый менеджер для работы через Интернет с файловой системой и подсистему для создания, редактирования, отображения реляционных данных.

Для организации выполнения WPS-сервисов необходимо учитывать их распределение в сети Интернет. Соответственно необходимо обеспечить инвентаризацию и поиск WPS-сервисов. Кроме того требуется более детальная спецификация входных и выходных параметров, чем это определено стандартом WPS. Детальная спецификация позволит упростить пользовательский интерфейс и проводить верификацию входных данных еще до запуска WPS-сервиса. Инвентаризацию и поиск WPS-сервисов реализует разработанный в рамках проекта каталог. Каталог WPS-сервисов разработан в виде модуля системы управления контентом Calipso. Регистрация WPS-сервиса в каталоге происходит в несколько этапов. На первом этапе пользователь вводит: название сервиса, его описание, данные для обращения к WPS-службе. На втором этапе каталог запрашивает, в соответствии со стандартом WPS, по введенному адресу метаданные и отображает список имеющихся WPS-сервисов. После выбора нужного сервиса, выполняется запрос на получение метаданных о параметрах WPS-сервиса. На последнем этапе пользователь дополняет информацию о параметрах: используемый для ввода элемент управления и его свойства, пользовательское название параметра, поясняющий текст. Данная информация применяется для генерации поль-

зовательского интерфейса, верификации параметров и запуска WPS-сервиса. Поиск в каталоге может производиться по названию и описанию работы WPS-сервиса.

Для запуска пользователем WPS-сервиса генерируется форма ввода параметров. Для каждого параметра используется элемент управления, указанный при регистрации WPS-сервиса. Разработаны следующие элементы управления: `edit` – для ввода строковых значений; `number` – для ввода числовых значений; `checkbox` – для ввода булевых значений; `rectangle` – для указания экстенда (прямоугольной области на карте) и т.д. При запуске WPS-сервисов необходимо упростить указание данных из СХД и СУБД в качестве значений входных и выходных параметров. Поэтому разработаны элементы управления, которые работают с данными пользователя в СХД и СУБД: `file` – для выбора файла из СХД; `file_save` – для сохранения файла в СХД; `select_table` - для выбора таблицы из СУБД; `select_table_attr` - для выбора атрибута таблицы и т.д. При наличии на форме элементов управления `file`, `file_save`, `rectangle` создается карта, на которой отображаются данные, связанные с этими элементами. Набор элементов управления является расширяемым.

После ввода пользователем в форме Web-браузера данных среда должна обеспечить запуск WPS-сервисов и передачу, получение этих данных в соответствии со стандартом WPS и политиками регламентации доступа. В форме WPS-сервиса формируется список значений параметров и передается запрос подсистеме управления WPS-сервисами на выполнение. Подсистема управления сервисами производит обработку параметров для их передачи в соответствии со стандартом WPS, регистрирует экземпляр выполняемого WPS-сервиса, запускает WPS-сервис, получает и сохраняет результаты работы. В зависимости от типа параметров обрабатываются значения следующим образом:

1. Файл. Если передаваемый параметр является файлом, расположенным в директории пользователя в СХД, то необходима передача данного файла WPS-сервису. В соответствии со стандартом WPS-сервису передаются URL адрес файла, используя который, WPS-сервис должен по протоколу HTTP скачать файл. Файл должен находиться в открытом доступе. Так как данные пользователей могут содержать информацию ограниченного доступа, то любой файл, хранимый в СХД, по умолчанию не может быть свободно доступным. Таким образом, необходимо одновременно защитить и предоставить WPS-сервисам доступ к хранимым в СХД файлам. Механизм доступа и контроля обращений к файлам построен следующим образом – каждому экземпляру выполнения WPS-сервиса присваивается уникальный идентификатор. Если выполняемый WPS-сервис должен получить в качестве параметров файлы, то для каждого из файлов создается уникальная ссылка, привязываемая именно к выполняемому экземпляру. Таким образом, WPS-сервис может свободно скачать требуемый файл из СХД по сгенерированной ссылке определенное число раз, причем все ссылки, созданные для определенного экземпляра, уничтожаются как только метод завершит свою работу. Если файл является результатом работы WPS-сервиса, то подсистема выполнения сценариев загружает в СХД в соответствующую пользовательскую директорию по протоколу HTTP. URL-адреса файлов берутся из метаданных в ExecuteResponse документе, который формируется по окончании работы WPS-сервиса.

2. Экстент (прямоугольная область на карте). На клиентской части (браузере) получение экстента производится в формате WKT [1]. Для передачи WPS-сервису такого типа данных производится преобразование в соответствии со стандартом WPS.

3. Таблица базы данных PostgreSQL. При передаче в качестве параметра ссылки на таблицу формируется строка соединения

драйвера GDAL для непосредственного соединения WPS-сервиса с базой данных. В текущей версии таблицы PostgreSQL доступны только для локальных WPS-сервисов, но в дальнейшем будет реализация для регламентированного доступа к базе данных извне.

Интеграция WPS-сервисов осуществляется в виде сценариев, определяющих последовательность применения, передаваемые параметры и т.д.. Для разработки сценариев WPS-сервисов предлагается использовать язык JavaScript, где обращение к WPS-сервисам производится с помощью специальных функций. Использование языка JavaScript обладает рядом преимуществ, в частности JavaScript является полноценным языком программирования с поддержкой асинхронного выполнения и сохранением контекста, что очень важно, учитывая возможную длительность выполнения WPS-сервисов. Для интерпретации сценариев на языке JavaScript и непосредственного обращения к WPS-сервисам разработан специальный модуль, написанный на C++ с использованием JavaScript интерпретатора Google V8.

9. Разработанные WPS-сервисы

В рамках геопортала разработан ряд сервисов. Перечислим некоторые из них:

Сервис расчета плотности точечных объектов в ячейках регулярной сетки. На входе сервиса слой векторных объектов. На выходе количество объектов, находящихся в ячейках регулярной сетки. Пользователь может задать размер ячейки и область обработки. Сервис производит подсчет количество объектов в ячейках, а если задан атрибут семантики слоя входных данных, то производится суммирование значений указанного атрибута. Данный сервис используется, например, для расчета выбросов от точечных объектов.

Сервис расчета плотности линейных объектов в ячейках регулярной сетки. На входе сервиса слой линейных векторных объектов. На выходе общая длина участков линейных объектов, находящихся в ячейках регулярной сетки. Пользователь может задать размер ячейки и область обработки. Сервис для каждого линейного объекта производит трассировку и суммирование длин участков линейных объектов в ячейках, а если задан атрибут семантики слоя входных данных, то производится умножения длины участка объекта в ячейке на значение атрибута, а затем общее суммирование. Данный сервис используется для расчета выбросов дорожной сети.

Сервис изменения значений атрибутов. Данный сервис необходим для массового задания значений атрибута объектов, находящихся внутри полигонов. На входе сервиса название атрибута источника, слой источник полигональных объектов с желаемыми значениями атрибута и результирующий векторный файл (точечный или линейный), в котором значения атрибута меняются. Например, данный сервис можно использовать для задания характеристик печного отопления в определенных районах населенного пункта.

Сервис интерполяции точечных данных на ячейках регулярной сетки методом естественных соседей. На входе сервиса слой точечных векторных объектов в виде векторных файлов (поддерживаемых библиотекой GDAL) или таблицы PostgreSQL. На выходе интерполируемые значения в ячейках регулярной сетки, в формате GeoTIFF. Метод естественных соседей используется из библиотеки CGAL.

Сервисы построения карт рельефа, уклонов, экспозиции на основе данных радарной топографической съемки (SRTM) на

указанную территорию. Пользователь указывает экстенд желаемой территории и размер ячейки. Сервис автоматически определяет требуемые файлы данных SRTM и выполняет обработку. Результат работы сервисов сохраняется в виде набора файлов в формате GeoTIFF.

Сервис расчета нормализованного разностного вегетационного индекса (Normalized Difference Vegetation Index – NDVI) на основе мультиспектральных космических снимков. Производит расчет на основе данных пользователей, которые должны находиться в системе хранения данных геопортала. Результат сохраняется в формате GeoTIFF.

Сервисы алгебры GRID, которые позволяют производить умножение на число, сложение, вычитание GRID данных. С помощью данных сервисов, например, можно объединить набор GRID файлов.

Сервисы, обеспечивающие пространственный статистический анализ территорий. Вычислительное ядро сервиса составляют математические методы, реализованные в известной библиотеке AlgLib Free edition. В созданном сервисе реализованы возможности анализа пространственных данных через нахождение коэффициентов корреляции и регрессии. На вход для анализа поступают два файла формата GeoTIFF, описывающих различные характеристики одной территории. Для описания серии файлов используется специализированный файл в формате MTIFF. В этом случае серия GeoTIFF файлов позволяет описать динамику изменения некоторой характеристики в течение времени. Информация из GeoTIFF файлов экстрагируется и представляется в виде одномерных массивов данных с плавающей точкой. Между этими данными проводится анализ и поиск зависимостей.

Заключение

Комбинация WPS-сервисов и подсистем геопортала СХД, СУБД PostgreSQL позволяет упростить обработку пространственных данных в Интернет. Для передачи данных WPS-сервисам достаточно разместить их в геопортале с помощью подсистем ввода/редактирования данных, файлового менеджера и FTPS-сервера. Результаты работы WPS-сервисов размещаются также в геопортале. В дальнейшем каталог WPS-сервисов может выступать как общедоступная база сервисов, решающих задачи разных предметных областей. Возможность интеграции сервисов посредством специальных Javascript функций, делает геопортал удобной средой разработки новых методов поддержки междисциплинарных научных исследований.

Литература

1. OGC 05-007r7, OpenGIS® Web Processing Service / редактор: Peter Schut [Open Geospatial Consortium, Inc., 2007]. URL: <http://www.opengeospatial.org/standards/wps> (датаобращения: 21.10.2012).
2. OpenGIS Web Processing Service (WPS) Implementation Specification, v1.0.0. Release date: June 08, 2007. – URL: <http://www.opengeospatial.org/standards/wps> [15 февраля 2012]