

УДК 004.65: 004.62: 004.9

ПРОБЛЕМЫ ЗАГРУЗКИ ПРОСТРАНСТВЕННЫХ ДАННЫХ В ИНТЕГРИРОВАННУЮ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНУЮ СИСТЕМУ

Миронова А.И. студент гр. М-134, IV курс,
Харлампенков И.Е. (ИВТ СО РАН)

Научный руководитель: Гиниятуллина О.Л., к.т.н., доцент
Кемеровский государственный университет, ИВТ СО РАН
г. Кемерово

В настоящее время информационно-вычислительные системы переживают новый этап развития, который связан с публикацией данных в сети Интернет. В сети представлены проекты, которые решают конкретные производственные задачи с использованием пространственных данных в удаленном режиме, такие как веб-гис «Фаунистика», центр «Антистихия», серия проектов НСВФ [1-3]. Проекты осуществляют сбор, анализ и обработку геоданных, а также введение банка данных в свою область применения.

Имеется ряд программных продуктов для работы с пространственными данными (Google Earth, GeoMedia, ArcGIS, gvSIG, QGIS, GRASS). В настоящее время существует общий подход к созданию распределенных информационно-вычислительных систем для обработки геоданных.

В КФ ИВТ СО РАН разработана и широко используется интегрированная информационно-вычислительная система динамической оценки экологического состояния угледобывающих районов, которая реализована в QGIS, с использованием объектно-реляционной СУБД PostgreSQL.

При этом довольно острым является вопрос загрузки пространственных данных в систему. Чаще всего пространственные данные хранятся в шейп-файлах. Из-за своей распространенности шейп-файл стал стандартным форматом для обмена данными между геоинформационными системами [4]. Известны стандартные загрузчики пространственных данных из шейп-файла в PostgreSQL, такие как shp2shp, ogr2ogr, QGIS Spit [5]. Загрузчики предназначены для конвертации векторных данных из одного формата в другой. Недостатком известных способов является то, что они не позволяют настроить параметры загрузки, выявить некорректные данные, проанализировать файл и выдать сообщение пользователю до извлечения данных.

В рамках интегрированной информационно-вычислительной системы динамической оценки экологического состояния угледобывающих районов в КФ ИВТ СО РАН разрабатывается подсистема загрузки пространственных данных, которая позволит реализовать указанные выше недостатки.

В результате анализа предметной области и существующих программных решений, сформулированы следующие требования к загрузчику:

- Создание и сохранение неограниченного количества подключений к БД;
- Выбор любого подключения из сохраненных;
- Проверка названия таблицы и колонок на русские символы, с возможностью их переименования до загрузки;
- Возможность добавления комментария к таблице;
- Возможность переименовать таблицу перед загрузкой в любом случае;
- Выявление некорректно заполненных строк пространственных данных и исключение их из таблицы;
- Возможность добавить строки в таблицу с таким же названием (при условии, если столбцы идентичны).

В качестве решения предлагается форма, которая базируется на предоставлении пользователю ряда настроек при загрузке файла, а также на выводе сообщений пользователю об ошибках и некорректных данных в файле с последующим предложением действий. Отличительная черта предлагаемого нами метода заключается в том, чтобы пользователь еще до загрузки данных имеет представление о том, как пространственные данные будут отображены в БД. Пользователь может контролировать процесс загрузки и иметь неограниченный список подключений, как к локальным серверам, так и к удаленным.

В качестве среды разработки загрузчика на языке программирования Java использовался NetBeans IDE, для построения и управления проектами - Apache Maven.

На рис 1 представлена UML-диаграмма активности, на которой показано разложение деятельности программы на ее составляющие. Под деятельностью понимается спецификация исполняемого поведения в виде координированного последовательного и параллельного выполнения подчинённых элементов — вложенных видов деятельности и отдельных действий, соединённых между собой потоками, которые идут от выходов одного узла к входам другого.

Как можно заметить, программа полностью ориентируется на ответы пользователя на возникающие спорные моменты. Реализованы все требуемые проверки. По окончании работы программы на экран выводится информация о том, сколько строк добавлено, а сколько было исключено из-за некорректно заполненной пространственной информации. Вся информация об ошибках исключенных строк хранится в отдельном текстовом файле, доступном пользователю.

Тестирование загрузчика осуществлялось на выборке из 57 shp-файлов. Эксперимент продемонстрировал успешную загрузку всех предложенных шейп-файлов. Из полученных результатов видно, что настроив параметры перед загрузкой можно избежать ряд трудоемкой работы по исправлению некорректных данных, загруженных в БД.

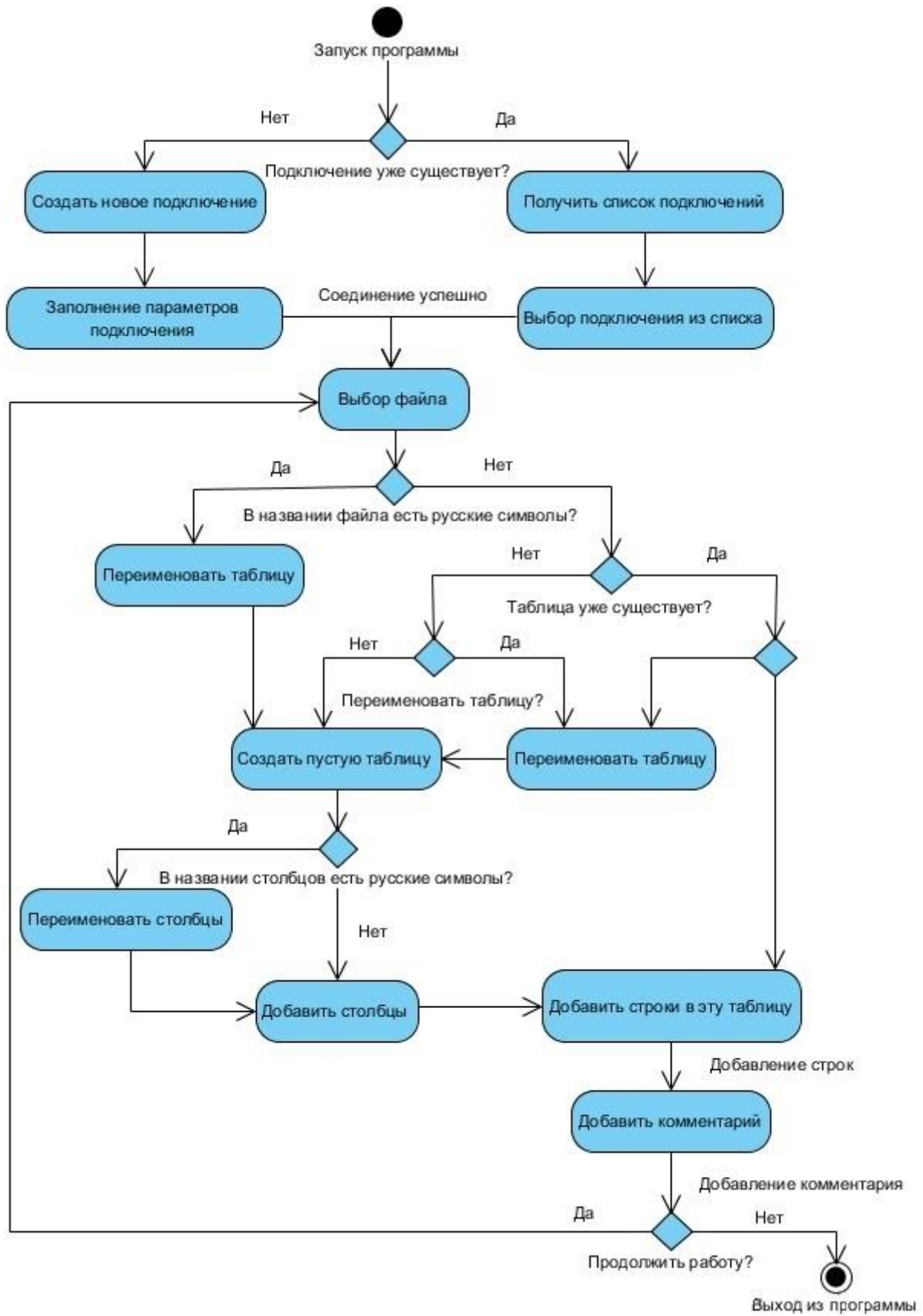


Рис 1. Диаграмма деятельности загрузчика

Таким образом, созданный загрузчик обладает большим преимуществом перед всеми доступными альтернативами. Безошибочно загружает пространственные данные и все их атрибуты в установленную БД. Исключает некорректно заполненные строки, что у альтернативных загрузчиков вызвало бы остановку загрузки и вывод ошибки.

Предлагаемый способ позволяет ускорить загрузку и редактирование пространственных данных, за счет настраиваемых параметров. Загрузчик может найти применение в любой сфере, где производится сбор, хранение и анализ пространственных данных.

Список литературы:

1. Всероссийский центр мониторинга и прогнозирования чрезвычайных ситуаций МЧС России (Центр "Антистихия") [Электронный ресурс] // URL: <http://www.mchs.gov.ru/document/89093> (дата обращения: 03.04.2017).
2. Веб-гис «Фаунистика» [Электронный ресурс] // URL: <http://rrrcn.ru/birdwatching/web-gis> (дата обращения: 03.04.2017).
3. Главные характеристики ЛВПЦ [Электронный ресурс] // URL: <http://hcvf.ru/ru/concept> (дата обращения: 03.04.2017).
4. Работа с векторными данными [Электронный ресурс] // URL: http://docs.qgis.org/2.0/ru/docs/user_manual/working_with_vector/index.html (дата обращения: 05.04.2017).
5. Руководство по PostGIS [Электронный ресурс] // URL: <http://gis-lab.info/docs/postgis/manual/> (дата обращения: 05.04.2017).