

РАЗРАБОТКА СПЕЦИАЛИЗИРОВАННОГО ПО ДЛЯ ОБРАБОТКИ ДАННЫХ ПО ЛЕСНЫМ МАССИВАМ ГРОДНЕНСКОЙ ОБЛАСТИ

А.И. Жвалевский, Н.И. Игнатенко

УО «Гродненский государственный университет им. Я. Купалы», факультет математики и информатики, специальность «Программное обеспечение информационных технологий», кафедра современных технологий программирования.

Научный руководитель – Л.В. Рудикова, кандидат физико-математических наук, доцент, заведующий кафедрой современных технологий программирования.

В статье рассматриваются основные подходы к построению и реализации архитектуры системы, предназначенной для обработки данных по лесным массивам Гродненской области, излагаются ключевые идеи, связанные с их анализом. Во введении описаны: поставленная проблема и цель работы. В основной части дается подробный отчет об особенностях проектирования и разработки компонентов системы и прогнозной модели. В заключении приводятся ключевые сводки по результатам работы.

Ключевые слова: лесовосстановление, веб-приложение, Предметно-ориентированное проектирование (DDD), глубокое обучение.

Введение. Сегодня все больше и больше внимания по всему миру уделяется различным проблемам экологии. Во многих лесных массивах ведется активная вырубка лесов и на местах вырубок осуществляется посадка искусственных насаждений, а естественное возобновление лесов имеет минимальную площадь. Примером здесь может послужить особо охраняемая природная территория (ООПТ) – «Гродненская пуща». Так как ООПТ создаются для сохранения и воспроизводства естественных сообществ (естественного биоразнообразия), то проблемы лесовосстановления и лесоразведения стоят перед ними особенно остро.

Лесовосстановление — это создание лесных культур на площадях, ранее покрытых лесом. Под лесоразведением понимают создание лесных культур на нелесных площадях. Лесоразведение проводят в основном на старопахотных землях, переданных в лесной фонд под облесение, в районах защитного лесоразведения и при лесной рекультивации земель.

Современных экологов крайне интересует целесообразность посадки искусственных насаждений, поскольку на начальном этапе своего роста геоботанические параметры, снятые с естественных особей и особей искусственного насаждения, заметно отличаются друг от друга. Из этого заурядного факта вытекает вполне очевидный вопрос, следует ли ожидать, что эти параметры сойдутся, и если сойдутся, то когда. Помимо вопросов лесовосстановления и лесоразведения работники рассадников сталкиваются с проблемой отсутствия специализированных систем для хранения и обработки информации, которые они регулярно собирают. Эти дилеммы дают нам понимание о значимости и эффективности использования таких систем.

Целью данной работы является разработка ПО для обработки данных по лесным массивам, которое позволит исследователю того или иного лесного массива упростить процесс записи и хранения информации о лесах и лесных породах, предоставит средства для ведения статистики с возможностью визуализации. Важной особенностью системы выступает возможность импорта данных из Excel-документов, которые многие предприятия используют для хранения информации, в базу данных приложения. Процесс записи данных и доступа к ним должен быть возможен, как с мобильного устройства (смартфона и планшета), так и с персонального компьютера. Особое внимание должно быть уделено поддержке работы приложения с мобильных устройств: возможность записи данных во время их снятия в «полевых условиях» является критическим фактором для эффективного использования. Данная система также станет фундаментом для последующего анализа данных, связанного с разработкой прогнозной модели по возрастной изменчивости посаженных и натуральных сосновых лесов для выявления временной точки (условий) совпадения параметров, т.е. прогнозируется, в каком возрасте данные леса по изученным параметрам не будут отличаться друг от друга.

Основная часть. Во время разработки было принято решение использовать Domain Driven Design, предметно-ориентированное проектирование. Такой подход позволяет наиболее правильно понять бизнес-процессы, выработать Ubiquitous Language (вездесущий язык) соответствующий предметной области. Такой подход позволит системе гибко реагировать на изменения в требованиях, но в то же время он требует всесторонней первоначальной проработки бизнес-процессов и предполагает аналитику, потом проектирование и лишь затем – разработку системы.

Система разделена на 3 контекста.

1. Контекст идентификации пользователя. Отвечает за регистрацию пользователя, вход в систему и контроль доступа к данным.

2. Контекст, отвечающий за хранение данных, получение, обновление и удаление информации о ООПТ.

3. Контекст, отвечающий за анализ данных, производит необходимые расчеты по требованию пользователя.

Контексты построены на основе 3-х уровневой архитектуры. Условно их можно разделить на следующие слои: Application, Domain и Infrastructure (рисунок 1).

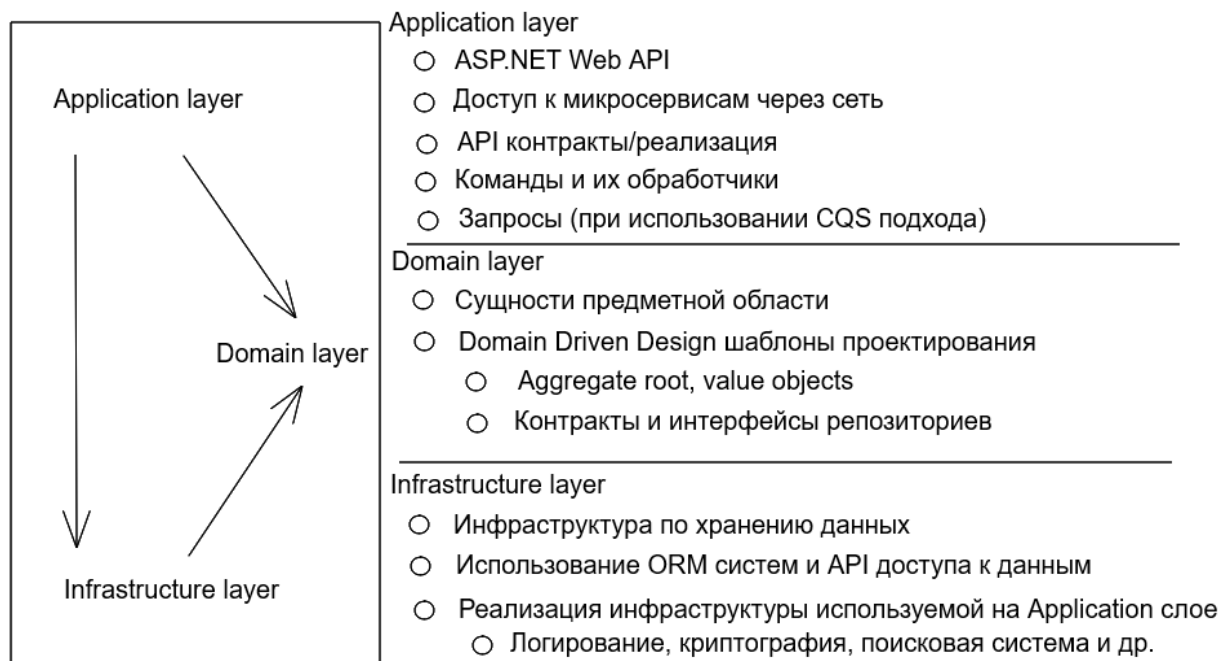


Рисунок 1 – Слои и связи между ними в Domain Driven Design

В рамках Гродненского ООПТ рассматривалась следующая иерархия сущностей. Имеется две группы лесов: естественного (натурального) и искусственного происхождения. В каждом лесу изучались возрастные группы (каждый лес разного возраста): I возрастная группа (1-20 лет), II возрастная группа (21-40 лет) и т.д. Для хвойных лесов (еловых, сосновых) – шаг 20 лет, для лиственных (дубовых, осиновых и т.д.) – шаг 20 лет. Также в каждом лесу изучались 4 яруса: деревья, молодые деревья, кусты и травы.

Во всех натуральных и посаженных лесах сняты экологические параметры: сквозистость (%), относительная освещенность (%), pH, гумус (мг/кг), P2O5 (фосфор) (мг/кг), K2O (калий) (мг/кг).

В соответствии со снимаемыми параметрами и глобальными требованиями была спроектирована следующая база данных (рисунок 2). Имеется сущность User (пользователь) хранящая основную информацию пользователя. У User есть Refresh Token (токен обновления), которые используются для создания токена доступа, позволяющего пользователю работать с системой. Также пользователь может создавать сущность Project, которая в свою очередь может содержать в себе несколько сущностей Forest (лес). Сущность Project может иметь несколько участников (User), каждый из них имеет разные права доступа, задаваемые создателем проекта (Owner). Сама сущность Forest может содержать несколько сущностей AgeClass (возрастная группа) и те в свою очередь имеют несколько Tier (ярусов). Также AgeClass содержит сущность EcologicalParameters (экологические параметры) и WeatherStatistic (статистика о погоде). Каждый Tier содержит определенные сущности Species (особи), которые соотносятся с сущностью SpeciesType (тип особи). Сущность Species может содержать несколько сущностей GeobotanicalParameter (геоботанические параметры).

Создание прогнозной модели выступает куда более неочевидной задачей. В данной работе она рассматривалась как нахождение экстремума (минимума или максимума) целевой функции в некоторой области конечномерного векторного пространства. Для решения этой проблемы было решено использовать машинное обучение, в частности, нейронные сети, потому как создание модели, использующей другую структуру, которая учитывает такое массивное количество параметров, представляется малореализуемым.

Прогнозная модель разделяется на 2 группы нейронных сетей: для насаждений естественного характера и для искусственных насаждений. В каждой группе отдельно обучены нейронные сети для каждого биологического параметра. Отсюда данные, поступающие на вход нейронной сети можно ограничить следующими факторами: возрастная группа, ярус, тип особи, год (для которого считается результат), экологические параметры (сквозистость (%), относительная освещенность (%), pH, гумус (мг/кг), P2O5 (фосфор) (мг/кг), K2O (калий) (мг/кг)). Таким образом, каждая отдельная нейронная сеть считает прогноз по одному биологическому параметру, характерному для определенного типа леса, яруса и особи при заданных экологических параметрах и годе. Такой подход позволяет децентрализовать модель, ускорить время

обучения и ее точность, позволив использовать разные методы и средства при проектировании отдельных нейронных сетей. На основе этого были выбраны следующие архитектуры: если биологический параметр - группа, то использовались softmax в качестве функции активации и cross-entropy в качестве функции ошибки на последнем слое. В противном случае, если параметр - числовая характеристика, то логистический сигмоида или ReLU выступают как активаторы и среднее квадратическое отклонение как ошибка.

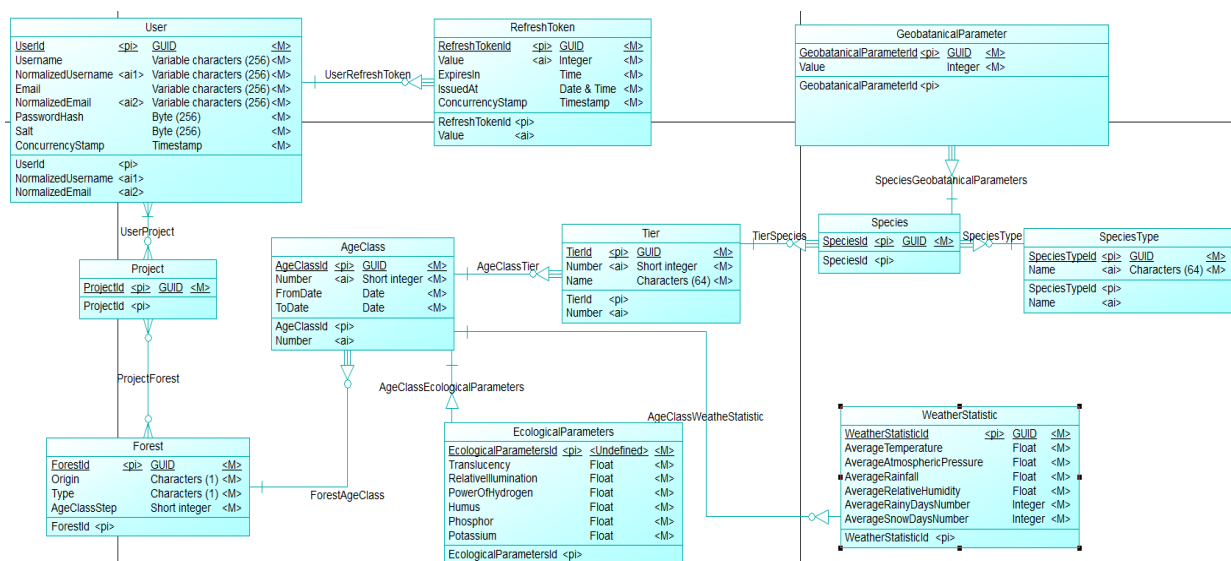


Рисунок 2 – Entity-Relationship схема базы данных

Используя разработанные нейронные сети, можно без труда спрогнозировать в каком возрасте данные леса по изученным параметрам не будут отличаться друг от друга или хотя бы частично сойдутся. К сожалению, объективность вышеописанных эвристических моделей все еще остается под вопросом.

Заключение. Результатом разработки является система, позволяющая биологам, изучающим ООП структурировано хранить свои данные, иметь к ним обширный доступ и регулярно получать результаты анализа данных. Также имеется возможность совместной работы нескольких специалистов, позволяющая разграничить возможности того или иного пользователя и упростить коммуникацию между ними. Особое внимание было уделено возможности перенести уже имеющиеся данные в систему, поэтому был разработан функционал по импорту данных из файлов.

На основе полученных данных были созданы прогнозные модели по возрастной изменчивости саженных и натуральных сосновых лесов для выявления временной точки (условий) совпадения параметров, т.е. прогнозируется, в каком возрасте данные леса по изученным параметрам не будут отличаться друг от друга.

Наша система может быть улучшена путем добавления возможности экспорта данных, для использования их в более специализированных программах, обеспечивая обратную совместимость данных. Наиболее важным улучшением будет добавление возможности по анализу данных определенной особи, отдельно от леса (брусника). Также целесообразно будет добавить больше факторов и данных влияющих на результаты обработки данных.

Список литературы

1. Николенко С., Кадури А., Архангельская Е. Глубокое обучение. – СПб.: Питер, 2018. – 480 с.: ил. – (Серия «Библиотека программиста»).
2. Дейт, К. Дж. Введение в системы баз данных, 8-е издание.: Пер. с англ. – М.: Издательский дом "Вильямс", 2005. – 1328 с.: ил. – Парал. тит. англ.
3. Предметно-ориентированное проектирование [электронный ресурс] \ Википедия. – Режим доступа: https://en.wikipedia.org/wiki/Domain-driven_design
4. Эрик Эванс. Предметно-ориентированное проектирование (DDD). Структуризация сложных программных систем – МСК.: Вильямс, 2010. – 448 с.