ОСОБЕННОСТИ РАЗРАБОТКИ МОДУЛЯ АНАЛИЗА ДАННЫХ ПЛАНИРОВАНИЯ ТРАНСПЛАНТАЦИЙ СТВОЛОВЫХ КЛЕТОК

FEATURES OF THE DEVELOPMENT OF THE DATA ANALYSIS MODULE FOR PLANNING STEM CELL TRANSPLANTS



УДК 004

Штефко Сергей Валерьевич, студент магистратуры, МИРЭА – Российский технологический университет, г. Москва

Shtefko Sergey Valerievich, Graduate Student, MIREA - Russian Technological University, Moscow, e-mail: sergei4009@yandex.ru

Аннотация

В данной статье рассматриваются особенности разработки модуля анализа данных планирования трансплантаций стволовых клеток для одной из информационных систем ФГБУ «НМИЦ ДГОИ им. Дмитрия Рогачева» Минздрава России. Были выбраны и описаны основные инструменты разработки модуля анализа данных, которые позволят в полной мере обеспечить все потребности в рамках анализа данных в системе планирования трансплантаций стволовых клеток.

Annotation

This article discusses the features of the development of a module for analyzing data for planning stem cell transplants for one of the information systems of the Dmitry Rogachev National Medical Research Center of Pediatric Hematology, Oncology and Immunology. The main tools for the development of the data analysis module were selected and described, which will fully meet all the needs for data analysis in the stem cell transplant planning system.

Ключевые слова: модуль анализа данных, Python, Dash, WSGI, Plotly, React, здравоохранение.
Keywords: data analysis module, Python, Dash, WSGI, Plotly, React, healthcare.

В условиях активного внедрения медицинских информационных систем в деятельность медицинских центров появляется проблема обработки и анализа огромных объемов данных, накапливающихся в данных системах. На рынке медицинских информационных систем присутствует множество систем со встроенными модулями анализа данных, но нередко возникает проблема внедрения новой системы в процессы деятельности медицинской организации из-за сопутствующих сложностей, возникающих при миграции с одной системы на другую. Данная проблема была выявлена в ФГБУ «НМИЦ ДГОИ им. Дмитрия Рогачева» Минздрава России. В данном случае требуется разработка собственного модуля анализа данных.

Для решения данной проблемы было принято решение о разработке модуля анализа данных в рамках существующей информационной системы.

Аналитический модуль — это компонент информационной системы, выполненный как ее часть или как внешнее интегрируемое ПО, который решает комплекс задач анализа данных, хранящихся и использующихся в системе.

Данный модуль расширяет возможности системы в плане работы с данными и эффективности их использования. Согласно концепции информационно-аналитического обеспечения системы, предоставляющие возможность обработки информации, проектируются и функционируют с учетом следующих аспектов [1]:

- извлечение из многих источников разнородных данных, представленных в различных форматах и приведение их единому формату и единой структуре;
- аккумулирование информации, создание информационных массивов данных, применение технологий индексации и поиска;

- организация предоставления пользователям необходимой информации, которая требуется для принятия решений, реализации конкретных мер или программных действий в сфере основной деятельности пользователей информационно-аналитической системы;
- представление информации и результатов анализа в упорядоченной форме для эффективного восприятия пользователями.

Поскольку доступ к информационной системе планирования трансплантаций реализован с помощью веб-интерфейса, то аналитический модуль будет реализован как связка серверной части (**back-end**), обеспечивающей обработку запросов и агрегацию данных и клиентской части (**front-end**), обеспечивающей взаимодействие пользователя с модулем анализа данных и отображением агрегированных данных из источников данных, которыми в данном случае являются базы данных МИС (Рисунок 1).

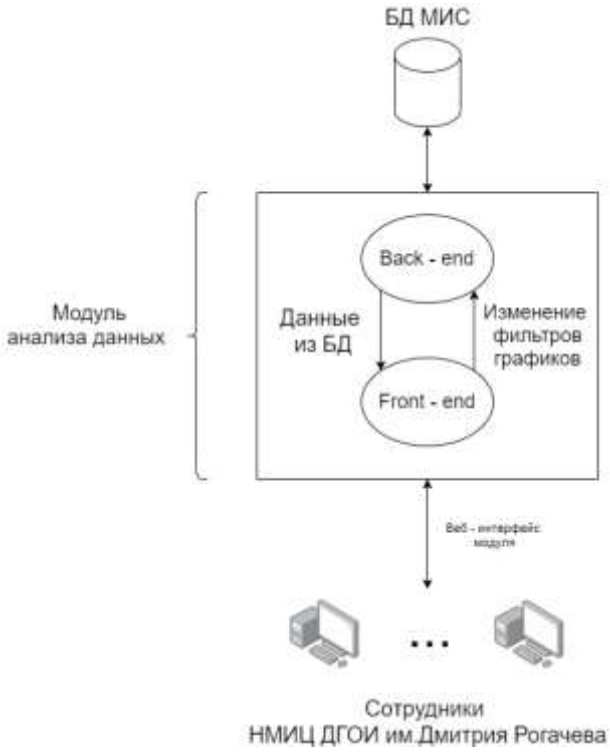

Рисунок 1 — Общая схема работы модуля анализа данных

Поскольку разрабатываемый модуль будет выполнять аналитические функции, необходимо это учесть на этапе выбора языка программирования.

Существует несколько языков программирования, которые активно используются для разработки аналитического приложения:

– Ruby;
– Python.

В рамках поставленной задачи был выбран язык Python, поскольку он обладает рядом преимуществ по сравнению с Ruby:

– в арсенале Python существуют интерпретаторы практически для всех операционных систем, существующих на данный момент, а значит с запуском программ, написанных на данном языке, не возникнет проблем;
– Python обладает огромным количеством библиотек, предназначенных для разных задач — визуализация данных,с татистика, машинное обучение и т.д.;
– в Python существует возможность подключить библиотеки, написанные на языке программирования C, что позволит получить различные преимущества данного языка без необходимости его изучения;
– наличие разнообразных источников информации о Python позволяет находить в кратчайшие сроки ответы на вопросы, возникающие при разработке [2].

Помимо использования языков программирования, целесообразно включить в средства разработки аналитического модуля так называемый framework.

Framework (иногда фре́ймво́рк; англицизм, неологизм от framework — остов, каркас, структура) — это программная платформа, определяющая структуру программной системы; программное обеспечение, облегчающее разработку и объединение разных компонентов большого программного проекта [3].

Учитывая функционал фреймворков, представленных на рынке целесообразно выбрать Dash, поскольку он наименее «перегружен» лишними

плагинами и обеспечит наименьшую нагрузку на сервер, что позволит увеличить производительность модуля, при этом не жертвуя его функционалом.

Dash это фреймворк состоящий всего из 3 компонентов:

1) Flask.
2) Jinja 2.
3) Plotly.

Flask — это простой маршрутизатор, который обрабатывает запросы и направляет ответы. Также он обеспечивает совместимость с WSGI для использования фреймворка с веб сервером.

Flask является фреймворком для веб-приложений, написанный на Python. Он разработан Армином Ронахером, который возглавляет международную группу энтузиастов Python под названием Россо.

Данный фреймворк основан на наборе таких инструментов, как Werkzeug WSGI и шаблонизаторе Jinja.

WSGI(Web-Server Gateway Interface) является потомком CGI (Common Gateway Interface). По мере разрастания интернета CGI также разрастался из-за поддержки огромного количества языков и из-за отсутствия других решений.

Преимуществами WSGI – сервера являются:

– WSGI-сервера были разработаны для одновременной обработки нескольких запросов, что немаловажно в рамках работы модуля анализа данных информационной системы, что позволит обеспечить одновременную работу нескольких специалистов;
– WSGI ускоряет разработку веб-приложений, написанных на языке Python;
– использование WSGI делает проект гибким, так как появляется возможность изменения компонентов веб-стека без изменения приложения, которое работает с WSGI[4].

Jinja 2 — это шаблонизатор который подставляет сгенерированные на сервере данные в сверстанную HTML страницу. Данная библиотека подобна

шаблонизатору Django, но предоставляет выражения, схожие по синтаксису с языком Python, обеспечивая исполнение шаблонов в песочнице.

Благодаря данному шаблонизатору появляется возможность обеспечить максимально быстрое создание дополнительных страниц (при возникшей необходимости) модуля анализа данных, в соответствии с базовым шаблоном,определенным для данного модуля.

Plotly является библиотекой с открытым исходным кодом, которая используется для создания интерактивных графиков. Данная библиотека использует декларативный подход, что существенно облегчает задачу визуализации. На практике это означает меньше усилий, затрачиваемых на создание фигуры, что позволяет программисту сосредоточиться на том, что представлять и как интерпретировать это [5].

С помощью Plotly становится возможным создать интерактивные визуализации данных, которые значительно ускорят процесс анализа данных в рамках данной системы планирования трансплантаций стволовых клеток.

Представленные выше инструменты разработки позволяют создать функциональный модуля анализа данных, который позволит в полной мере обеспечить все потребности в рамках анализа данных в системе планирования трансплантаций стволовых клеток.

Литература

1. Нестеров А.К. Информационно-аналитические системы // Энциклопедия Нестеровых: [Электронный ресурс] — Режим доступа: — http://odiplom.ru/lab/informacionno-analiticheskie-sistemy.html — Загл.с экрана. — Яз.рус.
2. ИПАП: Язык программирования Python: особенности и преимущества: [Электронный ресурс] — Режим доступа: https://ipap.ru/poleznye-stati/4-useful/yazyk-programmirovaniya-python-osobennosti-i-preimushchestva— Загл.с экрана. — Яз.рус.

3. Content Management Framework //Юнионпедия:[Электронный ресурс] — Режим доступа: https://ru.unionpedia.org/Content_Management_Framework — Загл.с экрана. — Яз.рус.
4. Введение в WSGI-серверы: Часть первая // Habr: [Электронный ресурс] — Режим доступа: https://habr.com/ru/post/426957/— Загл.с экрана. — Яз.рус.
5. Plotly: The front end for ML and data science models// Plotly : [Электронный ресурс] — Режим доступа: https://plotly.com/ — Загл.с экрана. — Яз.англ.

Literature

1. Nesterov A. K. Information and analytical systems // Encyclopedia of the Nesterovs: [Electronic resource] - Access mode: — http://odiplom.ru/lab/informacionno-analiticheskie-sistemy.html -Title from the screen. - Lang. rus.
2. IPAP: Python programming language: features and benefits: [Electronic resource] - Access mode: https://ipap.ru/poleznye-stati/4-useful/yazyk-programmirovaniya-python-osobennosti-i-preimushchestva-Title from the screen.- Lang. rus.
3. Content Management Structure //Unionpedia:[Electronic resource] - Access mode: https://ru.unionpedia.org/Content_Management_Framework -Title from the screen.- Lang. rus.
4. Introduction to WSGI servers: Part one / / Habr: [Electronic resource] - Access mode: https://habr.com/ru/post/426957/ - Title from the screen.- Lang. rus.
5. Plotly: front end for ML and data science models / / Plotly : [Electronic resource] - Access mode: https://plotly.com/ - Title from the screen.- Lang.eng.