

УДК 004.65:004.4:004.5;543.51

Коваленко А. Н.¹, Махнач Н. В.², Коваленко Е. И.², Сытова С. Н.¹, Черепица С. В.¹

ИНТЕГРАЦИЯ МОДУЛЕЙ МАСС-СПЕКТРОМЕТРИЧЕСКОГО ПОИСКА И ВИЗУАЛИЗАЦИИ ИЗОБРАЖЕНИЙ DICOM ФОРМАТА В LIMS ELAB

¹ Институт ядерных проблем Белгосуниверситета, Минск, Беларусь

² Белорусский государственный университет, Минск, Беларусь

Доклад посвящен демонстрации возможностей системы по интеграции различных видов данных и их представлений в рамках единого окружения, построенного на ядре информационной системы eLab. Для каждой из заявленных предметных областей (масс-спектрометрический поиск и визуализация изображений DICOM) решена проблема обеспечения представления в виде набора данных системы eLab.

Лабораторные информационные системы (ЛИС, LIMS) – важная составляющая для обеспечения хранения и обработки данных современных лабораторий. Обработка экспериментальных данных в масс-спектрометрии представляет собой сложную задачу, для решения которой требуются обширные знания многих пакетов программ, имеющих различные алгоритмы, требования к формату данных и пользовательские интерфейсы. Сбор, хранение и анализ данных медицинской визуализации – изображений компьютерной томографии (КТ), магнитной резонансной томографии (МРТ), ультразвуковых исследований (УЗИ) и т. д. – также требует применения специализированного программного обеспечения и работы с большими массивами данных. Реализация ЛИС в виде веб-приложения в настоящее время считается стандартным вариантом для лабораторий малого и среднего масштаба. Общей целью работы лаборатории аналитических исследований НИИ ЯП БГУ по разработке собственной ЛИС на базе ядра информационной системы eLab (Electronic Laboratory) [1] является разработка модулей ЛИС для хранения результатов масс-спектрометрии и идентификации неизвестных веществ с использованием масс-спектрометрических библиотек, а также медицинских изображений в виде баз данных SQL-сервера. Формат данных mzXML является общим форматом файлов на основе XML (eXtensible Markup Language – расширяемый язык разметки) для масс-спектрометрических данных протеомики, других биологических и химических направлений. Исходные проприетарные форматы файлов большинства производителей масс-спектрометрического оборудования могут быть преобразованы в открытый формат mzXML. Формат mzXML является международно-принятой спецификацией, что позволяет передавать данные, извлечённые из ЛИС, в другие системы и прикладные программы (например, ProteoWizard и др.). Отметим также промышленный стандарт для систем формирования медицинских изображений (КТ, МРТ, УЗИ) – это протокол и формат DICOM (Digital Imaging and Communications in Medicine).

Целью работы было: продемонстрировать возможности системы по интеграции различных видов данных и их представлений в рамках единого окружения, построенного на ядре системы eLab, а также обеспечить для масс-спектрометрического поиска и визуализации изображений DICOM представление в виде набора данных (dataset) системы eLab.

Система управления лабораторной информацией разработана таким образом, что различные предметные области (рабочие места) могут быть изолированы в виде отдельных мини-приложений (модулей), которые при этом отображаются в едином пользовательском интерфейсе, формируются из элементов стандартной библиотеки ядра системы, пользуются общими службами системы, не влияют друг на друга, так как пользуются различными базами данных, которые могут находиться на физически разных серверах. Также особенностью интерфейса системы является то, что объектом манипуляций пользователя является один набор данных (dataset), который представляет собой таблицу с типизованными колонками. Над таким набором данных пользователь может производить следующие действия:

- навигация (выбор страницы, диапазона записей);
- выбор колонок (только необходимая информация);

Секция 2. Прикладные проблемы информатики

- применение фильтров (указание какой текст должны содержать какие колонки);
- создать отчёт по результирующему набору данных (сформировать документ офисного пакета – электронную таблицу или текст).

Масс-спектрометрическая база данных (Модуль МС-поиска) включает в себя следующие компоненты: Таблицы Пиков, Спектров, Компонентов, Синонимов. Используемая схема БД описана в [2]. Результирующий набор данных – это список спектров, потому что поиск осуществляется по спектру и результатом является фильтрованный именно по этому критерию набор данных (рисунок 1).

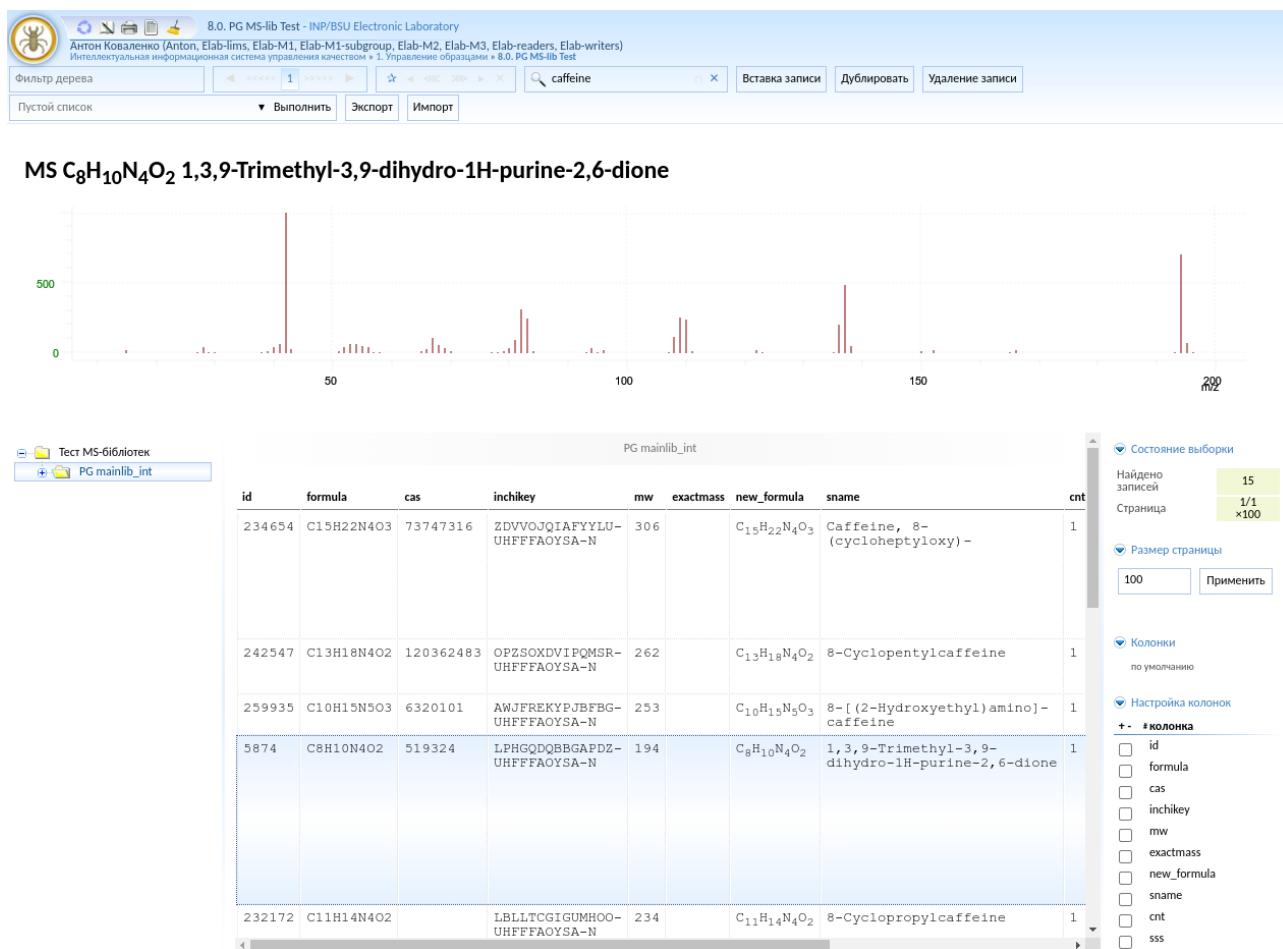


Рисунок 1 – Модуль масс-спектрометрического поиска системы eLab

Нами разработан конвертер файлов mzXML, предназначенный для формирования внутреннего представления масс-спектра (стандартная абстракция). С использованием данного конвертера поиск любого масс-спектра осуществляется по следующему алгоритму.

1. Получить файл «сырых» экспериментальных данных в любом поддерживаемом формате через пользовательский интерфейс (кнопка Загрузить) или через вызов системного API (<http://server.elab/modules/mslib/?method=upload>).
2. Выполнить преобразование данных закрытого формата в открытый mzXML (который выбран нами в качестве внутреннего представления) с помощью вызова файлового конвертера `/opt/unichrom/ucc input_file.raw output_file.mzxml`.
3. Выполнить XML-разбор файла `output_file.mzxml`.
4. Извлечь нужный скан (масс-спектр).
5. Передать спектр в текстовом виде (mz,int;...mz,int) во временную таблицу на SQL-сервере (SPECTRA_TMP).
6. Вызвать на SQL-сервере хранимую процедуру MSSEARCH(max_hit_cout).
7. Получить из хранимой процедуры список ID кандидатов спектров, сортированный по

Секция 2. Прикладные проблемы информатики

максимальному подобию.

8. Применить фильтр из полученных ID к списку масс-спектров и отобразить результаты поиска.

Задачи по реализации модуля DICOM поиска были следующие:

- обеспечить предварительную визуализацию данных МРТ, КТ исследований, хранящихся на свободно-распространяемом цифровом архиве медицинских изображений (picture archiving and communication systems – PACS) Orthanc;
- организовать просмотр и поиск с минимальной модификацией конфигурации хранилища данных.

Стандартная установка PACS Orthanc имеет собственный Web-интерфейс, который в ряде случаев не удовлетворяет требованиями клиник, но весьма удобен для администраторов систем. В связи с тем, что хранилище данных Orthanc (PostgreSQL) представляет собой денормализованную реляционную базу данных, имеется возможность подключиться к базам данных напрямую и обеспечить собственное представление данных. Схема базы данных опубликована и её анализ не представляет никаких проблем.

Результат работы реализованного алгоритма представлен на рисунке 2.

The screenshot shows the Orthanc web interface. On the left is a file tree for 'Сервер DICOM ort01'. The main area displays a table of search results with columns: internalid, resourcetype, publicid, parentid, and img. Three rows are visible, each with a corresponding MRI brain scan image. On the right, there are controls for search results, including 'Состояние выборки' (3 found records), 'Размер страницы' (100), 'Настройка колонок' (columns: internalid, resourcetype, publicid, parentid, img), and 'Сортировка' (sorted by internalid).

internalid	resourcetype	publicid	parentid	img
14	3	18534e3f-be89e32b-83dfb41b-c5569b33-a4d6e1e3	13	
15	3	59fccd7c-b192ef57-84b494b6-dc18d845-845f34ea	13	
16	3	9acba157-6d88298f-6f8f1c00-8b82a16f-470487b1	13	

Рисунок 2 – Модуль просмотра архива сервера PASC Orthanc

Таким образом, в работе выполнена интеграция различных видов данных и их представлений в единое окружение на базе ядра системы eLab.

Список литературы

1. Сытова С.Н., Дунец А.П., Коваленко А.Н. и др. Информационная система eLab для аккредитованных испытательных лабораторий на основе свободного программного обеспечения // Информатика. – 2017, № 3. – С.49–61.
2. Махнач Н.В., Трофименков И.В. Разработка универсального программного интерфейса для анализа масс-спектрометрической информации // 78-я научная конференция студентов и аспирантов Белорусского государственного университета : материалы конф. В 3 ч. Ч. 1, Минск, 10–21 мая 2021 г. / Белорус. гос. ун-т ; редкол.: В. Г. Сафонов (гл. ред.) [и др.]. – Минск : БГУ, 2021. – С. 136-139.