

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической теории упругости и биомеханики

**Проектирование и реализация модуля предобработки данных для
системы поддержки принятия врачебных решений**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 442 группы

направления 09.03.03 Прикладная информатика

механико-математического факультета

Фетисова Максима Михайловича

Научный руководитель
к.ф.-м.н., доцент

подпись, дата

Л.В. Бессонов

Зав. кафедрой
д.ф.-м.н., профессор

подпись, дата

Л.Ю. Коссович

Саратов 2020

Введение. Информационные технологии – это множество методов, производственного процесса и программно-технического средства, объединенных в такую технологическую цепочку, которая обеспечивает хранение, сбор, обработку, отображение и передачу информации. Цель функционирования данной цепочки (информационной технологии) – это возможность снизить трудоемкость процессов использования информационных ресурсов и повышение их надежности и оперативности. Современные медицинские предприятия производят и накапливают огромные объемы данных. Смотря, насколько эффективно используется эта информация врачами, управляющими органами, руководителями и зависит качество медицинской помощи, уровень развития страны в целом и каждого ее территориального субъекта в частности, а также общий уровень жизни населения. Информационные технологии стали не заменимой составляющей здравоохранения.

Решения класса системы поддержки принятия врачебных решений (далее СППВР), предназначены для помощи врачам с принятием врачебных решений и содержат в себе множество инструментов для обработки медицинских данных и выдачи рекомендаций о лечении пациентов на их основе. Данный проект позволяет снизить нагрузку на специалистов за счет оперативной обработки. Система помогает оцифровать и наладить использование медицинских знаний: данных из региональных медицинских информационных систем, клинических рекомендаций Минздрава РФ по каждой нозологии, результатов актуальных исследований и примеров успешных случаев лечения пациентов с тяжелыми диагнозами. Благодаря этому врачи могут более оперативно принимать решения о назначении терапии пациентам и контролировать их соответствие принятым в российской медицинской практике стандартам. В связи с этим **актуальность данной работы** определена необходимостью частичной реализации задачи по созданию модуля предобработки данных для системы поддержки принятия врачебных решений.

Системы поддержки принятия врачебных решений поможет производить выборку среди накопленных медицинских данных и ускорит постановку диагноза с дальнейшим лечением.

В связи с этим **актуальность данной работы** определена необходимостью предобработки хранимых системой данных для решения задач прогнозирования результата хирургического лечения.

Целью выпускной квалификационной работы является частичная реализации задачи по созданию модуля предобработки данных СППВР для решения задачи прогнозирования результата хирургического лечения. Для достижения цели данной работы поставлены следующие задачи:

1. Анализ принципов и подходов к накоплению медицинских данных;
2. Анализ реализации хранилища медицинских данных СППВР;
3. Проектирование системы витрин и представлений для редукции реляционной структуры до плоской таблицы;
4. Реализация спроектированных витрин и представлений;
5. Реализация прототипа интерфейса модуля предобработки данных.

Объект исследования - фрагмент обезличенных данных базы данных "Медицинская", предоставленный лабораторией «Системы поддержки принятия врачебных решений» образовательного научно-исследовательского института наноструктур и биосистем. БД «Медицинская» представляет собой пациенто-ориентированный баз данных, разработка которой выполнялась в рамках проекта «Протез-1». Для исследовательских целей был взят усечённый до 50 записей (кейсов) фрагмент БД «Медицинская», содержащий обезличенные и урезанные описания клинических случаев травм позвоночника.

Научная новизна - разработан модуль частичной реализации системы поддержки принятия врачебных решений, способный выполнять выборку по пациентам из основной базы данных.

В первой главе работы рассмотрен анализ предметной области.

В первой главе были рассмотрены предоставленные медицинские данные для более детального понимания представленной БД «Медицинская».

Для этого была проанализирована структура хранения данных в медицинских информационных системах. Основной единицей (объектом или сущностью) накопления и хранения данных в медицинских информационных системах является Человек. Именно к человеку, так или иначе «привязываются» все данные, накапливаемые в информационной системе. В медицинских информационных системах данные о каждом человеке делятся на два класса: данные о человеке, как личности, и данные о человеке, как пациенте.

Исследованы основные теоретические понятия. Как амбулаторная карта, так и история болезни представляют собой набор определенным образом структурированных данных, имеющих иерархическую структуру представлены на рисунке 1.



Рисунок 1 — Схема иерархической структуры хранения данных в медицинских информационных системах

На сегодняшний день документооборот медицинских учреждений регламентирован Приказом «Об утверждении унифицированных форм

медицинской документации, используемых в медицинских организациях, оказывающих медицинскую помощь в амбулаторных условиях, и порядков по их заполнению» 15.12.2014 N 834н (ред. от 09.01.2018). Именно на его основе организуется накопление данных в большинстве медицинских информационных систем. При этом не учитывается такой факт, что утвержденные почти 30 лет назад учетно-отчетные формы были ориентированы исключительно на их ручную обработку и не предполагали использования каких-либо средств вычислительной техники и информационных систем. Использование этих форм является целесообразным и оправданным лишь при отсутствии в медицинском учреждении комплексных медицинских информационных систем. Именно для этих условий были разработаны и утверждены эти учетно-отчетные формы.

Причина такого подхода кроется в позиции организаторов здравоохранения, определяемой формой N 025/у приказа N 834н: «Медицинская карта пациента, получающего медицинскую помощь в амбулаторных условиях».

Учетная форма N 025/у "Медицинская карта пациента, получающего медицинскую помощь в амбулаторных условиях" (далее - Карта) является основным учетным медицинским документом медицинской организации (иной организации), оказывающей медицинскую помощь в амбулаторных условиях взрослому населению (далее - медицинская организация).

В БД использованы медицинские термины, для упрощения восприятия информации будут введены текстовые формулировки, на основе существующих кодификаций.

Вторая глава выпускной квалификационной работы посвящена проектированию модуля предобработки данных для системы поддержки принятия врачебных решений.

База данных «Медицинская» дает необходимые для статистического анализа данные. По результатам анализа можно выделить в данных значимые факторы, и уже на основе всех накопленных данных и результатах проведенного анализа формировать прогнозы для новых случаев. Необходимо локализовать те попытки автоматизации "канцелярских" или "архивных" работ, имеющихся в любом офисе, которые не имеют серьезного обоснования с точки зрения системного анализа офиса как предприятия и задач совершенствования его деятельности.

Взаимодействие «Модуля прогнозирования» с базами данных существенно отличается от работы других модулей, которые обмениваются с БД небольшими объемами данных по одной операции или одному пациенту. Для «Модуля прогнозирования» идет речь о работе с большими массивами данных по выбранным классам случаев. При этом для работы «Модуля прогнозирования» необходимо получить сразу все имеющиеся данные из всех полей.

Эта особенность обусловлена различием структуры данных, необходимых для накопления максимально полного объема сведений о случае заболевания (БД «Медицинская»), и структуры данных, необходимых для разработки прогностической модели.

Ключевое свойство структуры БД «Медицинская» – гибкость. Эта гибкость позволяет максимально полно накапливать данные, при необходимости, сопоставляя экземпляру данных одного типа (клинический случай, этап планирования, этап контроля и др.) потенциально неограниченное количество экземпляров данных другого типа (биохимический анализ крови, лучевое исследование, результат опроса и др.).

В то же время, для построения модели прогнозирования результатов лечения необходим набор троек характеристических векторов:

- вектор, характеризующий состояние пациента при поступлении, в том числе его характер повреждения и/или заболевания;
- вектор, характеризующий выбранный вариант лечения;
- вектор, характеризующий результат лечения.

Ближайшая аналогия для структуры БД «Медицинская» – это «дерево» с несбалансированной и нерегулярной структурой, ER-диаграмма БД «Медицинской» представлена на рисунке 2.

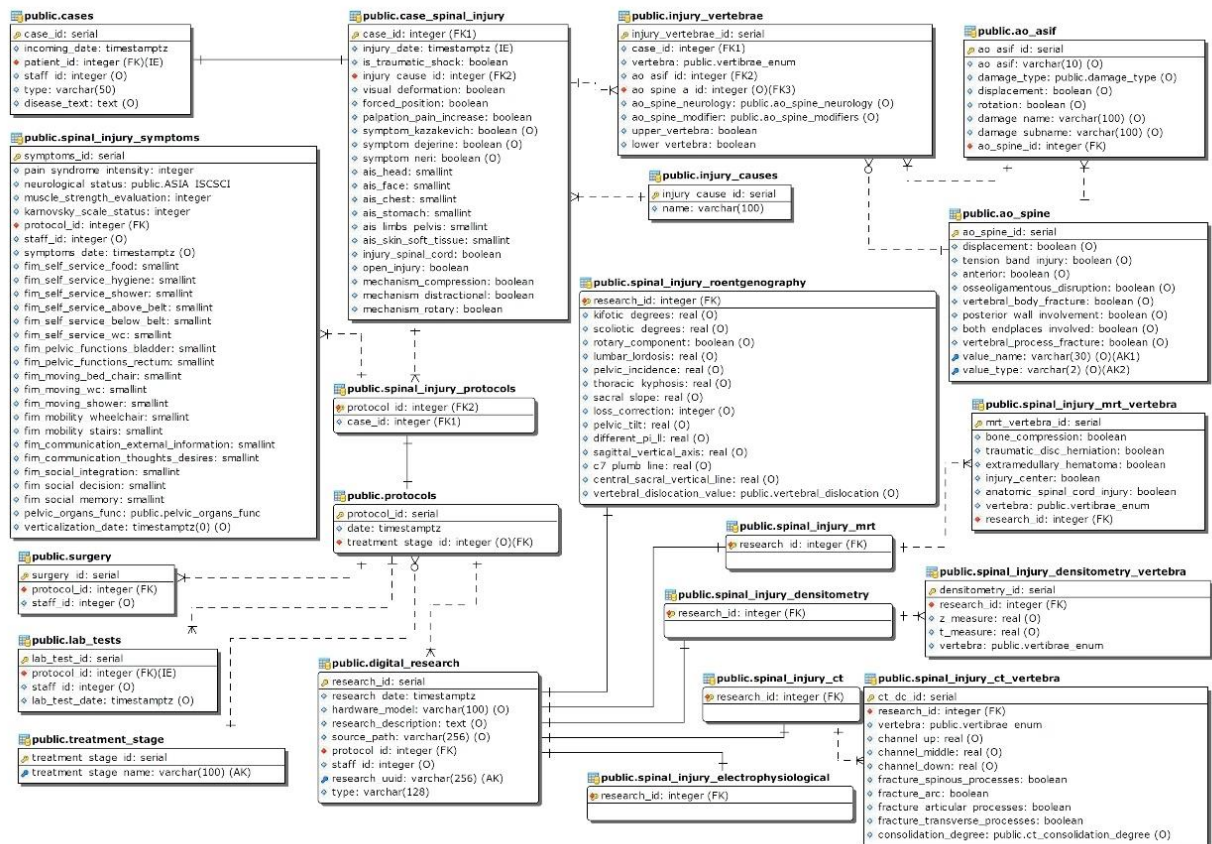


Рисунок 2 – Структура основных таблиц БД «Медицинская» для травм позвоночника

Витрины данных представляют собой срез БД «Медицинская», состоящий из узконаправленных тематических данных. При этом, в отличие от структуры БД «Медицинская», витрина может быть даже денормализована. Таким образом, открывается возможность широко применять механизмы агрегации и прочие механизмы, «портящие» исходные данные для нужд

решаемой задачи. Эти действия не несут негативных последствий, т.к. исходные данные БД «Медицинская» остаются неизменными. Технически, витрина – это таблица, заполнение которой данными происходит по некоторому алгоритму, посредством регламентных процедур. Данные, которыми заполняется витрина, являются результатом обработки исходных данных. Т.к. исходные данные остаются незатронутыми, витрина может быть безболезненно опорожнена в любой момент времени и сформирована заново – «пересчитана». На текущей стадии реализации проекта, учитывая небольшой по меркам теории баз данных накопленный объём информации БД «Медицинская», вместо витрин-таблиц можно использовать витрины-представления. Т.е. витрины могут быть реализованы как именованные запросы. В этом случае их содержимое будет вычисляться «налету», в момент обращения к витрине. Этот подход экономит время на реализацию витрин, однако он применим только в тех случаях, когда алгоритм формирования витрины может быть реализован средствами DML SQL, т.е. не требует полноты по Тьюрингу от языка реализации.

Для большей прозрачности архитектуры трансформации применяется декомпозиция: строятся цепочки служебных витрин, каждая из которых отвечает за отдельную ветку схемы сведения «древовидной» структуры данных к «плоской». Цепочка подразумевает, что витрина более высокого уровня использует данные витрины более низкого уровня пример приведен на рисунке 3 , а самые низкоуровневые витрины берут данные из таблиц БД «Медицинская». Этот приём применяется для таких видов исследований как опросники и для включения исследований в витрины протоколов (поступление, контроль 3 день, контроль 3 месяц, контроль 6 месяц, контроль 12 месяц).

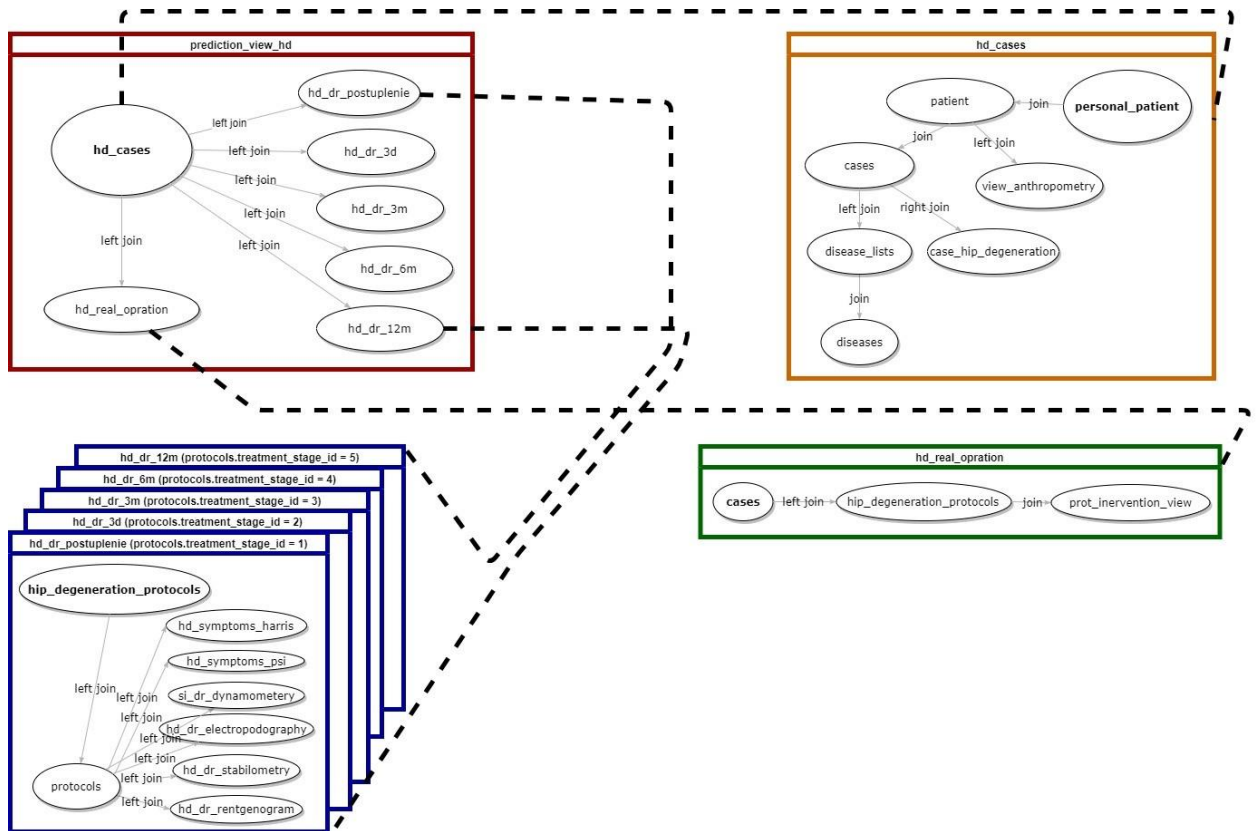


Рисунок 3 — Структура служебных витрин для витрины prediction_view_hd

Прототипирование интерфейса пользователя это одна из необходимых задач для разработки приложения. На данный момент существует достаточно много удобных сервисов. Такие сервисы чаще всего используют веб-разработчики и веб-дизайнеры для создания удобных и красивых проектов.

Для решения поставленной задачи было выбрано программное обеспечение Axure RP для создания прототипов и спецификаций веб-сайтов и приложений.

С помощью данной программы был создан графический интерфейс модуля предобработки данных для системы поддержки принятия врачебных решений. Пример интерфейса представлен на рисунке 4.

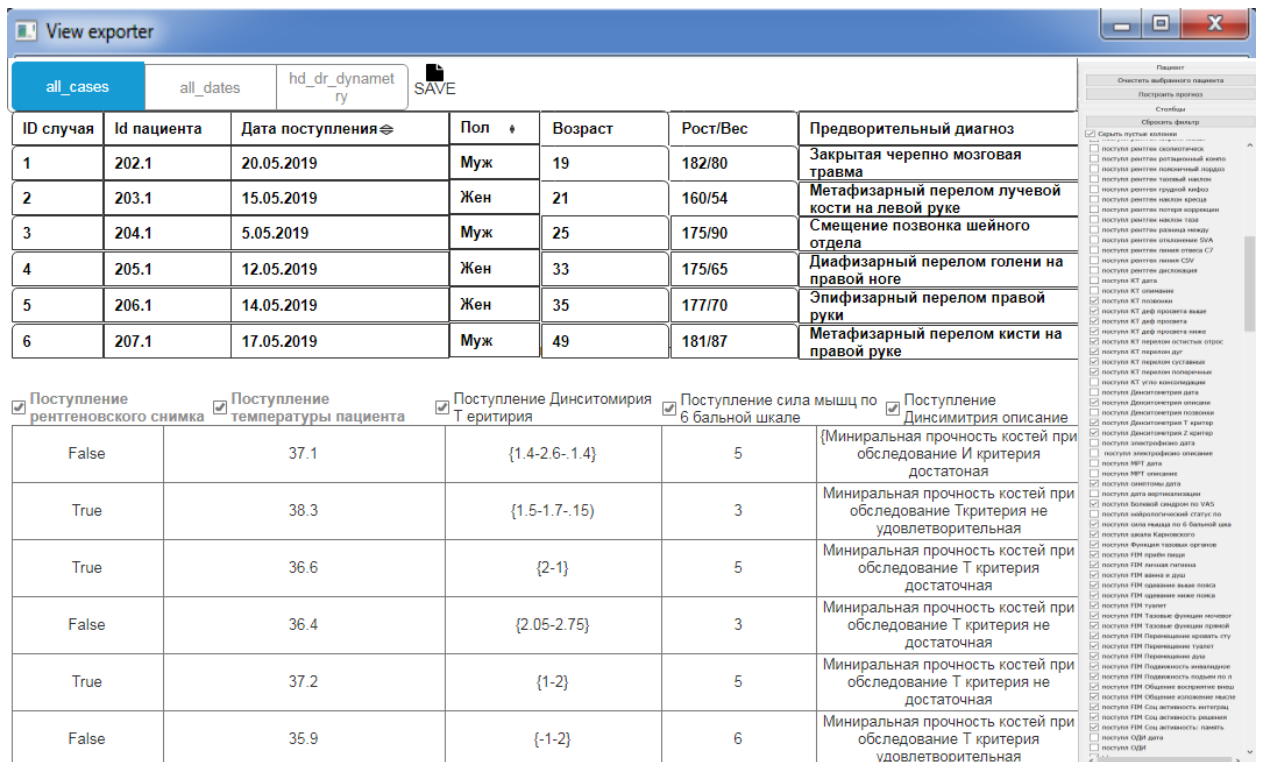


Рисунок 4— Пример графического интерфейса модуля предобработки данных для системы поддержки принятия врачебных решений

Третья глава выпускной квалификационной работы содержит реализацию разработанного модуля предобработки данных.

Витрины данных представляют собой срез БД «Медицинская», состоящий из узконаправленных тематических данных. При этом, в отличие от структуры БД «Медицинская», витрина может быть даже денормализована. Таким образом, открывается возможность широко применять механизмы агрегации и прочие механизмы, «портящие» исходные данные для нужд решаемой задачи. Эти действия не несут негативных последствий, т.к. исходные данные БД «Медицинская» остаются неизменными. Технически, витрина – это таблица, заполнение которой данными происходит по некоторому алгоритму, посредством регламентных процедур. Данные, которыми заполняется витрина, являются результатом обработки исходных данных. Т.к. исходные данные остаются незатронутыми, витрина может быть безболезненно опорожнена в любой момент времени и сформирована заново – «пересчитана». Пример структуры данных витрины all_cases приведен на рисунке 5



Рисунок 5 — Структура данных витрины all_cases

С помощью использованных методов был создан графический интерфейс для модуля предобработки данных для системы поддержки принятия врачебных решений. Интерфейс представлен на рисунке 6.

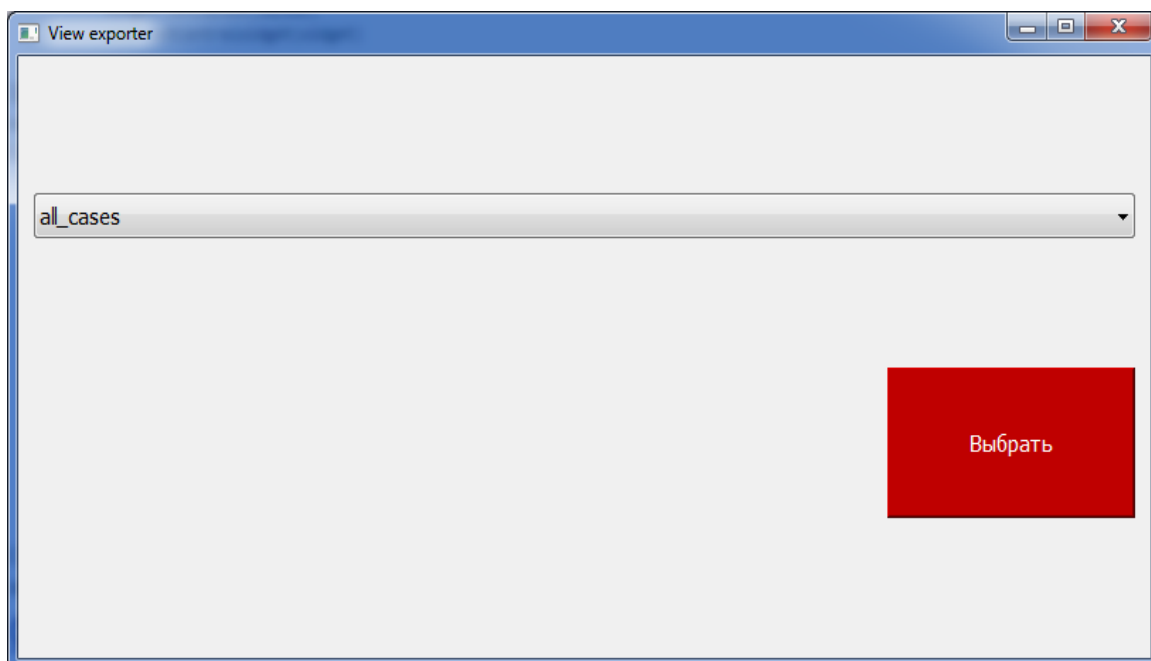


Рисунок 6 — Выбор экспортируемой view

На рисунке 7 представлены все таблицы, находящиеся в витрине all_cases

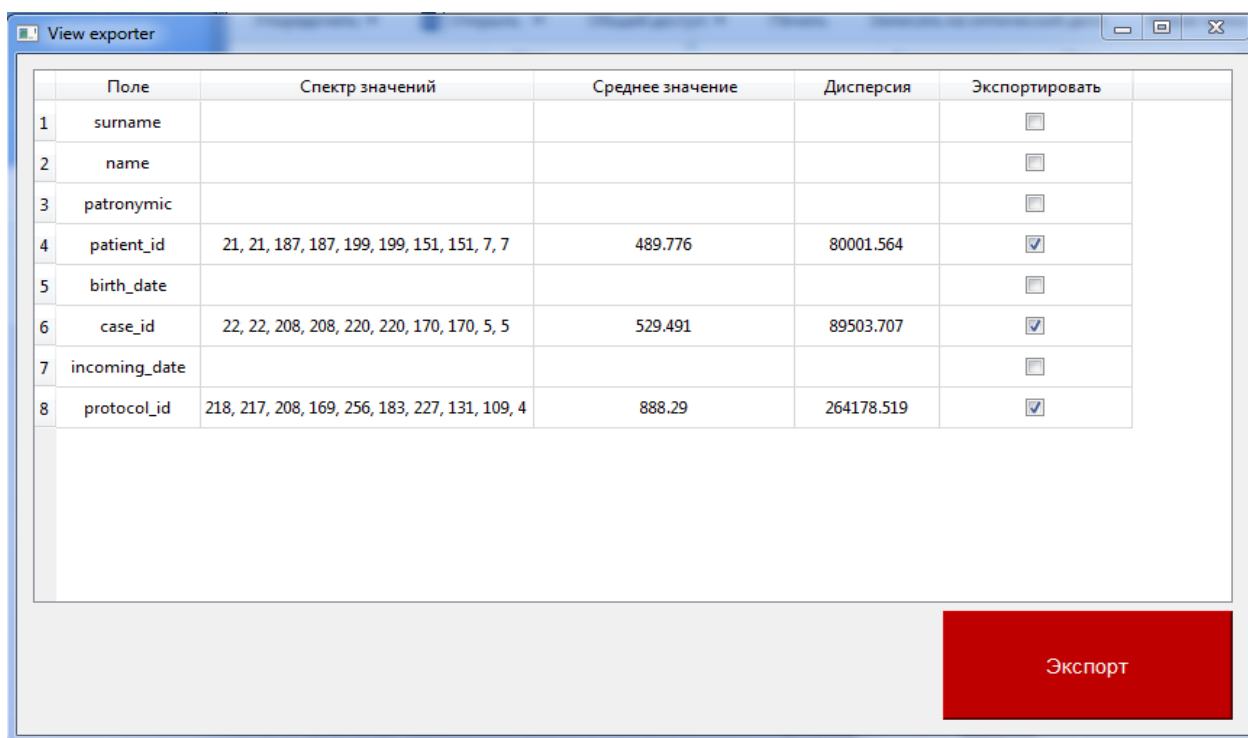


Рисунок 7 — Выбор необходимых таблиц для прогнозирования

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Зарубина, Т.В. Перспективы использования систематизированной номенклатуры медицинских терминов в России / Врач и информационные технологии. /Т.В. Зарубина, Е.С. Пашкина. – 2012. – № 4. – С. 4–14.
- 2 Зарубина, Т.В. Анализ описания расстройств сознания в SNOMED CT // Информационные технологии в медицине./Т.В. Зарубина, Е.С. Пашкина. – М., – 2012. – С. 30–41.
- 3 Пашкина, Е.С. О систематизированной номенклатуре медицинских терминов SNOMED CT вопросы полноты, аудита, сравнения, соответствия онтологическим стандартам. Врач и информационные технологии. /Е.С. Пашкина. – 2013. – № 2. – С. 71–78.
- 4 Пашкина, Е.С. SNOMED CT и проблемы терминологических систем // Врач и информационные технологии. / Е.С. Пашкина, Т.В. Зарубина. – 2013. – № 1. – С. 54–62.
- 5 Oliphant, T.E. Python for scientific computing. Computing in Science & Engineering. / T.E. Oliphant. – 2007. – Т. 9, № 3. – С. 10–20.
- 6 Rossum, V.G. Interactively testing remote servers using the Python programming language / G.V. Rossum, D. J. Boer. – CWi Quarterly. – 1991. – Т. 4. – №. 4. – С. 283-303.
- 7 Sanner, M.F. Python: a programming language for software integration and development /M.F. Sanner. – J Mol Graph Model. – 1999. – Т. 17. – №. 1. – С. 57-61.
- 8 Гагарина, Л.Г. Разработка и эксплуатация автоматизированных информационных систем. /Л.Г. Гагарина, Д.В. Киселёв. – М., – 2007. – С. 274–298.

- 9 Парфенов, Ю.П. Масштабируемое и отказоустойчивое dbass хранилище из линейки продуктов POSTGESQL. Программные системы и вычислительные методы. /Ю.П. Парфенов, Д.А. Девятериков. – 2014. – №. 1. – С. 125–130.
- 10 Akishina, E. Status of the geometry database for the CBM experiment / Akishina, CEUR Workshop Proceedings. – 2017. – С. 144–154.