

Төлен Г. Б.

Диодтың квазистатикалық үлгісінің видеоимпульспен әсер еткендегі қателіктері.

Түйіндемесі. Элементтің вольт-амперлі және вольт-фарадты сипаттамаларын беретін квазистатикалық өлшеу тәсілінің видеоимпульсті сигналмен әсер еткендегі жұмысы қарастырылған. Диодқа видеоимпульсті сигналмен әсер еткенде квазистатикалық үлгіде методикалық қателік бар екендігі дәлелденді. Себебі, идеалдандырылған p - n -өткелінің үлгісі диффузиялық зарядтың кешігуін ескермейді. Сондықтан, диффузиялық токтың кешігуін ескеретін үлгі жасау керек екендігі ұсынылды.

Кілтті сөздер: p - n -өткелі, квазистатикалық тәсіл, видеоимпульсті сигнал, SPICE-үлгі, вольт-амперлі сипаттама, вольт-фарадты сипаттама, диффузиялық ток.

Tolen G. B.

Performances of a quasistatic diode model in video pulse impact.

Summary. The work of the quasi-static measurement method is considered, which allows obtaining current-voltage (IVC) and capacity-voltage (CVC) characteristics of an element, by acting on it with a short video pulse signal. In the course of the work, it was proved that there is a methodical error in the quasistatic model with a video pulse action on the diode, since this model of an idealized p - n junction does not take into account the delay of the diffusion charge. Therefore, it was suggested to create a diode model that will take into account the delay of the diffusion current.

Key words: p - n junction, quasistatic method, video pulse signal, SPICE-model, current-voltage characteristic, capacitance-voltage characteristic, diffusion current.

УДК 004.65

М.Г. Бахытжан, А.А. Беделбаев

(КазНУ им. Аль-Фараби, Алматы, Республика Казахстан)

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ АРХИВНЫХ ДАННЫХ В СИСТЕМЕ ХРАНЕНИЯ ДОКУМЕНТОВ КАЗНУ НА БАЗЕ SQL SERVER FOR TOOLS 2012 КОМПАНИИ MICROSOFT

Аннотация: Рассмотрены методы интеллектуального анализа архивных данных и возможность их применения КазНУ на примере системы хранения документов КазНУ на базе SQL Server Tools. Приводятся математическая (основанная на нейросетевом подходе) и информационная (на языке UML) модели системы и решается при помощи современных алгоритмов кластеризации (дерево решений, логистическая регрессия, нейронные сети). В работе приводятся описания и оценка качества построенных моделей класс. Описанная в статье методика по ее применению позволяет значительно формализовать анализ данных и сократить затрачиваемые на него ресурсы. Разработанная система позволяет повысить уровень эффективности применения интеллектуального анализа архивных данных по определенным критериям и категориям.

Ключевые слова: интеллектуальный анализ данных, машинное обучение, классификации, кластеризация, система хранения документов, математическая статистика, кластерный анализ, метод ближайшего соседа.

В связи с совершенствованием технологий записи и хранения данных на людей обрушились колоссальные потоки информационной руды в самых различных областях. Деятельность любого предприятия (коммерческого, производственного, медицинского, научного и т.д.) теперь сопровождается регистрацией и записью всех подробностей его деятельности. Большие данные привели к взрывному росту популярности более широких методов интеллектуального анализа архивных данных.

Принципы интеллектуального анализа данных известны в течение многих лет, но с появлением больших данных они получили еще более широкое распространение. При работе с большими наборами данных уже недостаточно относительно простой и прямолинейной статистики. Имея около миллиона подробных записей необходимо знать не только информацию о месте нахождения нескольких тысяч записей, но также необходимо понимать принадлежит ли эта категория записей к определенной группе. Данные требования создали сложный процесс интеллектуальному анализу данных. Для решения задач требуется анализ+данных, в котором строится модель для описания информации, а также в заключение создается результирующий отчет.

Сегодня доступен совершенно новый спектр инструментов и систем, включая комбинированные системы хранения и обработки данных. Интеллектуальный анализ данных связан с поиском в данных скрытых нетривиальных и полезных закономерностей, позволяющих получить

новые знания об исследуемых данных. Известные статистические методы покрывают лишь часть нужд по обработке данных, и для их использования необходимо иметь четкое представление об искомых закономерностях. В такой ситуации методы интеллектуального анализа данных приобретают особую актуальность. Интеллектуальный анализ данных проник повсюду, вошел в стандартный набор функциональных возможностей информационной инфраструктуры многих отраслей.

В процессе интеллектуального анализа архивных данных выделяются несколько этапов: 1) понимание и формулировка задачи анализа; 2) подготовка данных для автоматизированного анализа; 3) применение методов интеллектуального анализа данных и построение моделей; 4) проверка построенных моделей; 5) интерпретация моделей человеком. В статье рассматриваем интеллектуальный анализ для архивных данных, где одним из первых этапов анализа является подготовка данных. Подготовка данных — применяется к области интеллектуального анализа данных, чтобы объединить предсказанные кластеризации, или от такой же модели для различных обучающих данных. Он также используется для преодоления внутренней нестабильности результатов при применении комплекса моделей для относительно небольших наборов данных. В нашем случае, мы готовим, именно, архивные данные, рассматриваем и изучаем их.

Данные для электронного архива, мы можем создать на основе шаблонов, путем переноса каталогов и файлов с локального и сетевого диска, ввод документов со сканера OCR версия. Или же можем, сразу воспользоваться услугами готовых электронных документооборотов, в нашем исследовании мы воспользуемся системой электронного документооборота Documentolog компании ТОО “Documentolog”.

Затем идет Учет. Индексация документов по информационным полям. Реквизиты жестко настраиваются и не могут быть изменены. Объединение существующих данных. Эти данные могут находиться в различных базах и на разных серверах. В этом случае можно использовать возможности служб SQLServerIntegrationServices. Наиболее простым из всех подходов часто служит опора на базы данных SQL. SQL (и соответствующая структура таблицы) хорошо понятен, но структуру и формат информации нельзя игнорировать полностью. Например, при изучении поведения пользователей по данным о продажах в модели данных SQL (и интеллектуального анализа данных в целом) существуют два основных формата, которые можно использовать: транзакционный и поведенческо-демографический.

При работе с InfoSphere Warehouse создание поведенческо-демографической модели в целях анализа данных о покупателях для понимания моделей их поведения предусматривает использование исходных данных SQL, основанных на информации о транзакциях, и известных параметров покупателей с организацией этой информации в заранее определенную табличную структуру. Затем InfoSphere Warehouse может использовать эту информацию для интеллектуального анализа данных методом кластеризации и классификации с целью получения нужного результата. Демографические данные о покупателях и данные о транзакциях можно скомбинировать, а затем преобразовать в формат, который допускает анализ определенных данных.

Во время анализа требований к системе основное внимание уделялось выяснению того, что должно быть сделано, вне зависимости от того, как это сделать. На этапе разработки системы решается вопрос, как реализовать решения, принятые на этапе анализа. Сначала разрабатывается общая структура (архитектура) системы. Архитектура системы определяет ее разбиение на модули, задает контекст, в рамках которого принимаются проектные решения на следующих этапах разработки. На рисунке 1 подробно показана общая архитектура системы электронного архива.

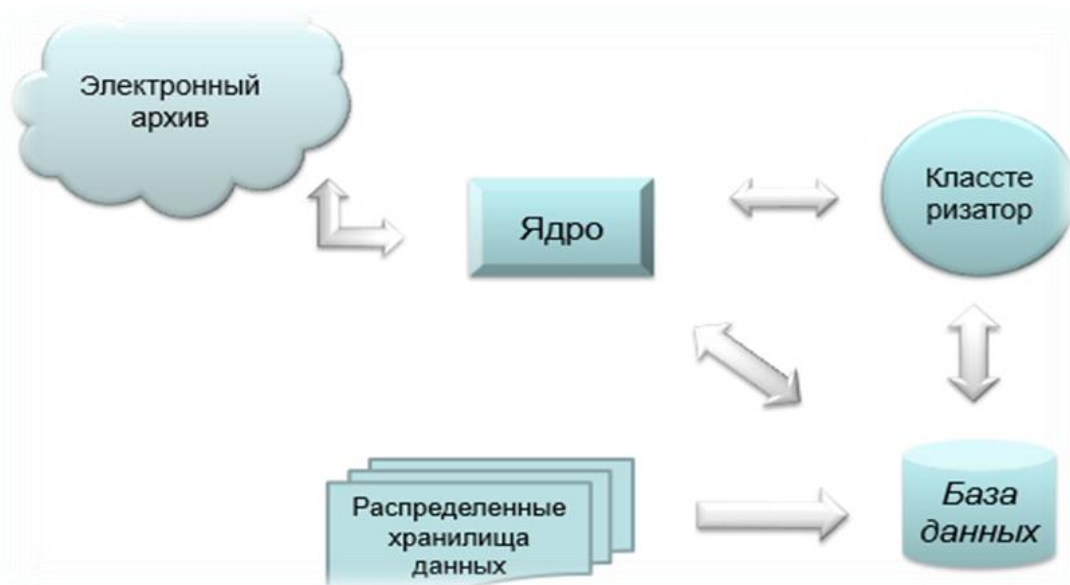


Рис. 1. Общая архитектура системы электронного архива

В исследовании мы берем место хранения данных КазНУ, как распределенные хранилища данных. При большом объеме данных и количестве пользователей мощности одного сервера может быть недостаточно. В этом случае нагрузка может быть перераспределена между несколькими файл-серверами и сервером базы данных. Распределение документов по нескольким хранилищам позволит сбалансировать нагрузку на сервера, сеть, увеличить максимальный объем хранимых данных, повысить отказоустойчивость системы в целом и снизить стоимость хранения. Хранилище данных SQL полностью интегрировано со службами SQL Server Analysis Services, Integration Services и Reporting Services.

Проектирование системы выполнялось с использованием унифицированного языка моделирования UML. Описание архитектуры системы приводится в нотации UML, и включает в себя описание диаграмм вариантов, диаграмму последовательностей, компонентной диаграммы, диаграммы класса и деятельности. Предполагается работа одного пользователя, который может вводить и модифицировать исходные данные (с использованием табличного процессора Ms-Excel).

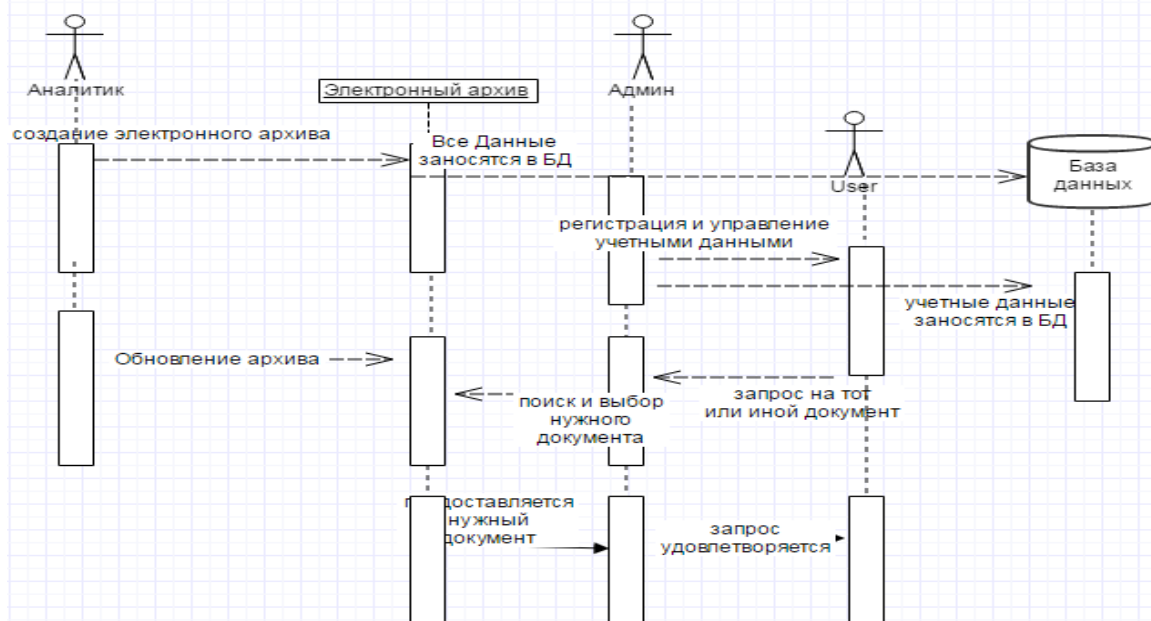


Рис. 2. Диаграмма последовательностей электронного архива

• Физико-математические науки

Диаграмма последовательностей отображает взаимодействие объектов в динамике, описывающей поведенческие аспекты системы, но рассматривает взаимодействие объектов во времени. На рисунке 2 на диаграмме последовательностей показан процесс последовательных действий со стороны аналитика, админа и пользователя в электронном архиве.

Наша система представляет собой совокупность моделей логического и физического уровней, которые должны быть согласованы между собой. Компоненты связываются через зависимости, когда соединяется требуемый интерфейс одного компонента с имеющимся интерфейсом другого компонента. Таким образом иллюстрируются отношения клиент-источник между двумя компонентами. В языке UML для физического представления моделей систем используются диаграммы компонентов. Такая диаграмма компонентов показана на рисунке 3.

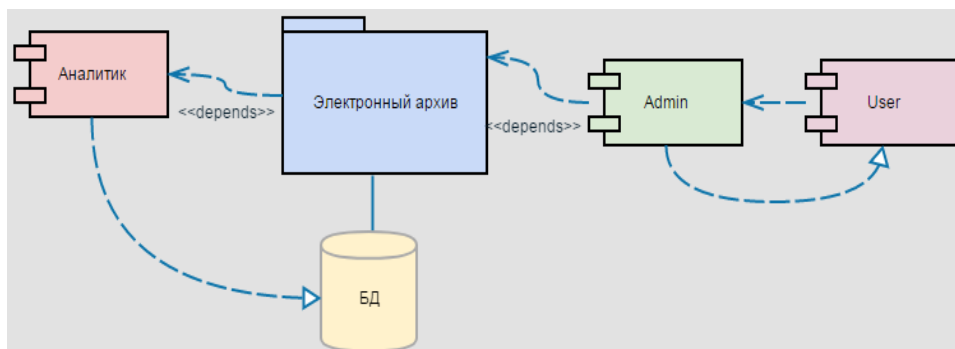


Рис. 3. Диаграмма компонентов UML электронного архива

Также, подробно описываем структуру нашего электронного архива, с помощью диаграммы классов (рисунок 4). Они являются одной из форм статического описания системы с точки зрения ее проектирования, показывая ее структуру.

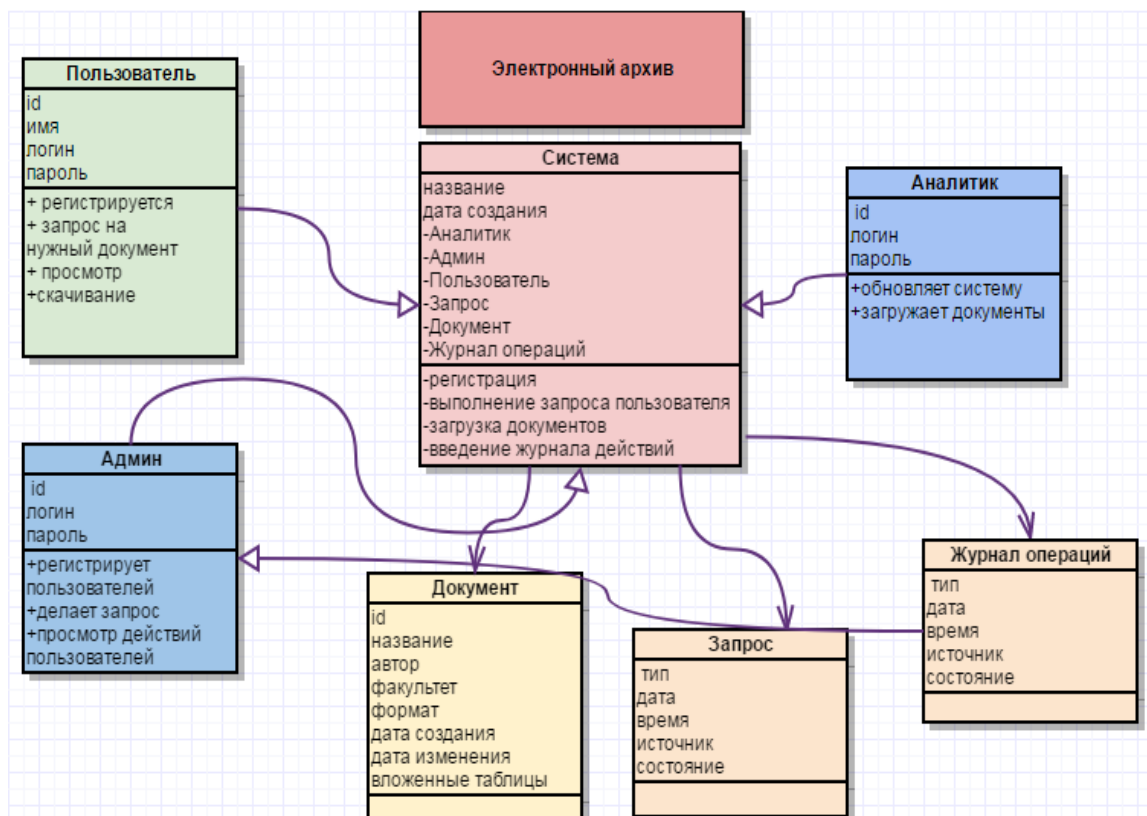


Рис. 4. Диаграмма классов

И чтобы показать жизненный цикл всех процессов в электронном архиве мы используем диаграмму деятельности (рисунок 5). Именно на диаграмме деятельности представлены переходы потока управления от одной деятельности к другой. Это, по сути, разновидность диаграммы состояний, где все или большая часть состояний являются некоторыми деятельностями, а все или большая часть переходов срабатывают при завершении определенной деятельности и позволяют перейти к выполнению следующей. Она может быть присоединена к любому элементу модели, имеющему динамическое поведение. В нашей диаграмме деятельности, показывается процесс запроса пользователя на поиск, скачивание или загрузку нужных документов в электронном архиве, и все эти процессы и запросы у нас делаются с помощью нашего администратора.

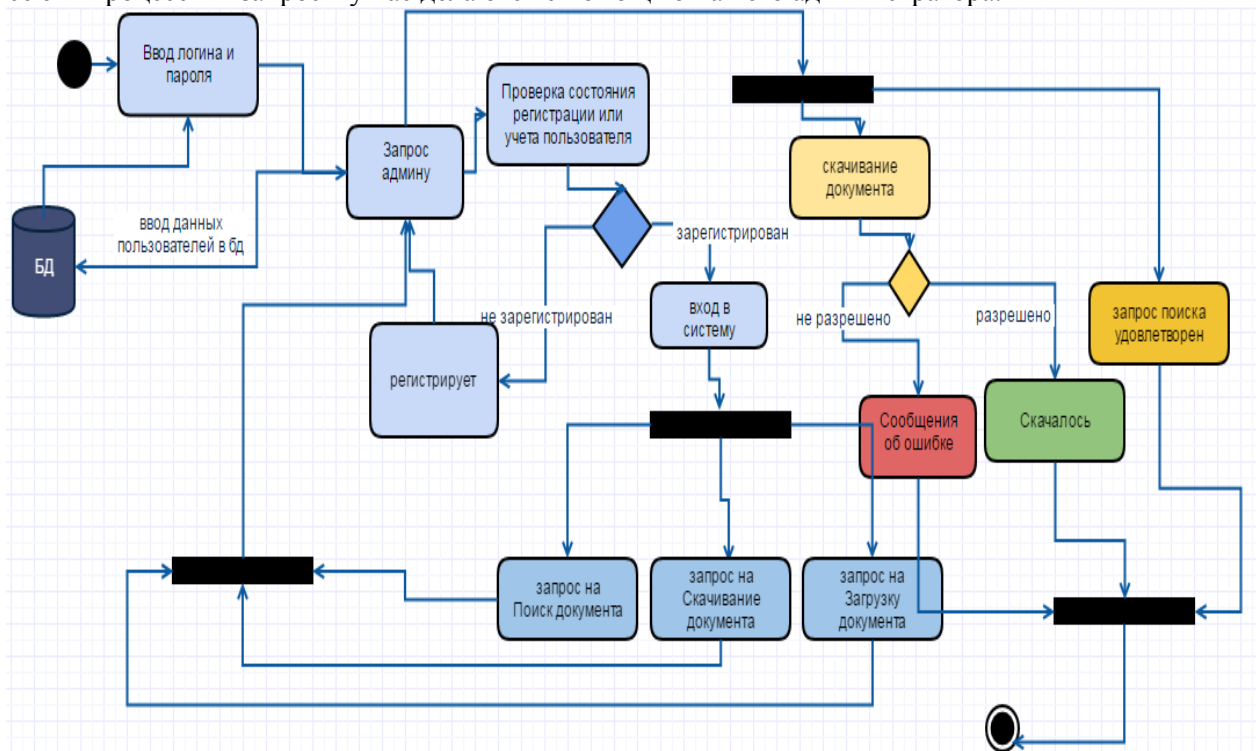


Рис. 5. Диаграмма деятельности

Дальше, идет изучение наших данных, согласно интеллектуальному анализу данных. Мы изучаем наши данные с помощью алгоритма кластеризации. Кластеризация — это задача более сложная, особенность кластеризации заключается в том, что классы объектов изначально не predetermined. Результатом кластеризации является разбиение объектов на группы. С помощью кластеризации средства интеллектуального анализа самостоятельно выделяют различные однородные группы данных.

Алгоритм кластеризации можно задать с помощью параметра *CLUSTERING_METHOD*. По умолчанию используется масштабируемая максимизация ожидания.

CLUSTERING_METHOD. Указывает метод кластеризации, используемый алгоритмом иерархического кластерного анализа.

- 1) Каждый объект — кластер.
- 2) Выбирает два кластера, расположенных ближе всего друг к другу и объединяет их.
- 3) Повторяет шаг 2. . . пока не объединит все кластеры.

Строим дендрограмму (рисунок 6) для остановки объединения.

Также используем алгоритм кластеризации последовательностей, который находит самые распространенные последовательности и выполняет кластеризацию для поиска идентичных последовательностей.

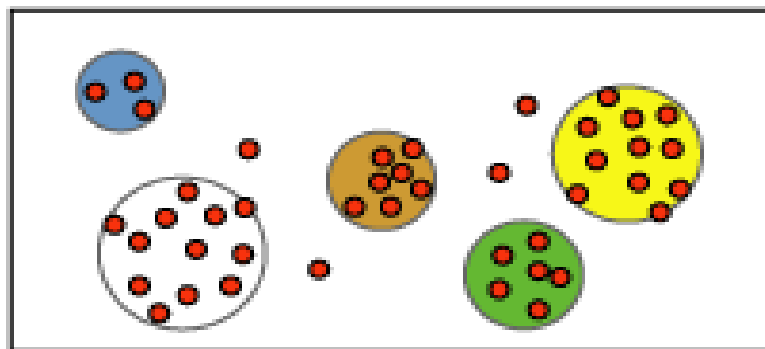


Рис. 6. Дендрограмма метода кластеризации

Далее для кластерного анализа используем метод ближайшего соседа. Метод ближайшего соседа (The nearest-neighbor method) классифицирует данные на основе аналогичных в предыдущих периодах. Аудиторы могут использовать этот подход для определения интересующего их документа и «попросить» систему выполнить поиск похожих элементов.

Следующий этап анализа наших данных является построение моделей. Знания, полученные при выполнении предыдущего этапа, помогут определить и создать модели. Разработчик определяет столбцы данных, которые должны быть использованы, путем создания структуры интеллектуального анализа данных. Структура интеллектуального анализа связана с источником данных, но не содержит никаких данных до обработки. Во время обработки структуры интеллектуального анализа службы Analysis Services создают статистические выражения и другие статистические данные, которые могут использоваться в анализе. Эти данные могут использоваться любой моделью интеллектуального анализа данных, которая основана на этой структуре. Модель интеллектуального анализа данных (рисунок 7) перед обработкой структуры и модели является просто контейнером, который задает столбцы, используемые для входных данных, прогнозируемый атрибут и параметры, управляющие алгоритмом обработки данных. Новую модель можно определить с помощью мастера интеллектуального анализа данных в среде SQL Server Data Tools или с помощью языка расширений интеллектуального анализа данных.

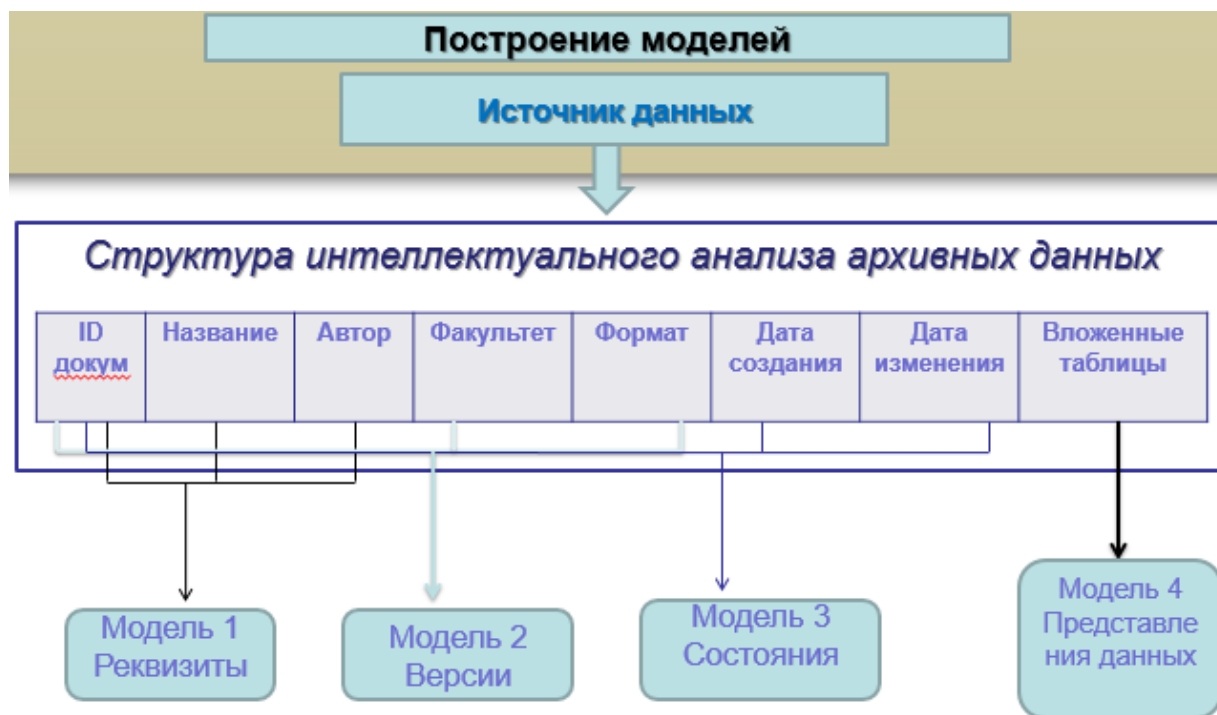


Рис. 7. Структура моделей интеллектуального анализа архивных данных

Службы AnalysisServices предоставляют конструктор создания модели.

Эти модели можно определить с помощью мастера интеллектуального анализа данных в среде BI DevStudio или с помощью языка *DMX*. Модели созданы, чтобы для сравнения результатов и выбора наилучшей пути поиска.

В MS SQLServer 2008 службы AnalysisServices предоставляют средства, упрощающие разделение данных.

Конечным этапом нашего интеллектуального анализа является развертывание. *Конечный пользователь*, в ответ на сформированный *запрос*, будет получать результаты анализа в виде отчета. При формировании отчетов о результатах проведенного анализа могут использоваться возможности службы SQLServer ReportingServices.

ЛИТЕРАТУРА

- [1] Чубукова И. А. Data Mining: учебное пособие. — Лаборатория знаний, 2006.
- [2] Кречетов Н.. Продукты для интеллектуального анализа данных. — Рынок программных средств, № 14–15, 1997
- [3] Дюк В., Самойленко А. Data Mining. Издательский дом "Питер". СПб, 2001.
- [4] Пескова, О.В. Методы автоматической классификации текстовых электронных документов / О.В. Пескова // Научно-техническая информация. Серия 2: Информационные процессы и системы. – 2006.
- [5] Певченко С. С. Методы интеллектуального анализа данных // Молодой ученый. — 2015. — №13.

Бахытжан М.Ғ., Беделбаев А.Ә.

Microsoft компаниясының SQL Server for tools 2012 негізінде ҚазҰУ құжаттар сақтау жүйесінде мұрағатты деректерді интеллектуалды анализдеу

Түйіндемe: Бұл мақалада мұрағаттық деректертерді интеллектуалды талдау әдістері және оларды ҚазҰУ құжаттарды сақтау жүйесі мысалында, SQL Server-Tools базасында ҚазҰУ қолдану мүмкіндігі қарастырылған. Математикалық (нейрожүйелік тәсілде негізделген) және ақпараттық (UML тіліндегі) жүйенің модельдері келтіріледі және қазіргі заманғы кластерлеу алгоритмдері (шешімдер ағашы, логистикалық регрессия, нейрондық желілер) көмегімен шешіледі. Осы еңбекте салынған модельдер класын сипаттау және сапасын бағалау келтіріледі. Мақалада сипатталған әдістемелер оны қолдануы бойынша деректерді талдауын айтарлықтай рәсімдеуге және оған жұмсалынатын ресурстарды қысқартуға мүмкіндік береді. Өзірленген жүйе анықталған өлшемдер мен санаттар бойынша мұрағаттық деректертердің интеллектуалды талдауының қолдану тиімділігінің деңгейін арттыруға мүмкіндік береді.

Түйін сөздер: деректерді интеллектуалды талдау, машиналық оқыту, топтастыру, кластерлеу, құжаттарды сақтау жүйесі, математикалық статистика, кластерлік талдау, жакын көрші әдісі.

Bakhytzhan M.G., Bedelbaev A.A.

Intelligent analysis of archival data in KazNU documents storage system, on the basis of SQL Server for tools in 2012 of Microsoft Company

Summary: This article describes the methods of mining of archival data and their applicability by KazNU, in the example of KazNU documents storage system, on the basis of SQL Server for tools. There are given mathematical (based on neural network approach) and informational (in UML) of the system models and solved using modern clustering algorithms (decision tree, logistic regression, neural networks). The work presents the description and evaluation of the quality of the constructed models class. Described in the article the methodology for its application allows significantly to formalize the analysis of the data and to reduce resources. The developed system allows improving the effectiveness of mining archived data according to certain criteria and categories. Key words: data mining, machine learning, classification, clustering, documents storage system, mathematical statistics, cluster analysis, nearest neighbor method.