БАЗИ ДАНИХ, БАЗИ ЗНАНЬ ТА ІНЖЕНЕРІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

УДК 622:65.011.56:658.562.4:004.77.

Азарян А.А., Азарян В.А., Гриценко А.Н.
Криворожский национальный университет
Кайгородов Р.А., Мирошник Д.Ю.
ООО «РУДПРОМГЕОФИЗИКА»

ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ КОНТРОЛЯ КАЧЕСТВА ПРИ ДОБЫЧЕ ЖЕЛЕЗОРУДНОГО СЫРЬЯ В УСЛОВИЯХ КАРЬЕРОВ

Целью данной работы является обоснование проекта по решению проблемы связанной с контролем качества и оперативного управления добычей железорудного сырья на всех стадиях технологической цепи от забоя до отгрузки товарной руды потребителям, обзор основных положений проекта таких как актуальность, постановка задачи, функциональная схема и ее проблемные зоны, алгоритм работы системы.

Метою даної роботи є обґрунтування проекту за рішенням проблеми пов'язаної з контролем якості й оперативного керування здобиччю залізорудної сировини на всіх стадіях технологічного ланцюга від забою до відвантаження товарної руди споживачам, огляд основних положень проекту таких як актуальність, постановка задачі, функціональна схема і її проблемні зони, алгоритм роботи системи.

The aim of this work is to substantiation of the project to address the issue associated with quality control and operational management of mining of iron ore at all stages of the chain from slaughter to the consumers of commodity ore, An overview of the project such as relevance, problem statement, functional diagram and problem areas, algorithm of the system.

Ключевые слова:контроль качества железорудного сырья, управление добычей и переработкой, автоматизированная система контроля качества.

.

Введение

Горно-металлургическая отрасль представляет собой стратегическую отрасль экономики страны, роль которой состоит в обеспечении сырьем производства высокотехнологичной и наукоемкой конечной продукции (машиностроения, стройиндустрии, авиационной, космической и оборонной промышленности).[1] Сегодня, добыча сырья и производство конечной продукции требует больших физических и материальных затрат. С каждым днем потребности в новом оборудовании, устройствах и автоматизированных системах управления растут с геометрической прогрессией.

На данный момент, ООО «РУДПРОМГЕОФИЗИКА» совместно с проблемной лабораторией Криворожского национального университета (КНУ) и ООО «НПП «Тетра» активно принимает участие в развитии горно-металлургической отрасли.

Основное направление научной деятельности предприятия:

- оперативный контроль и управление качеством минерального сырья при добыче, переработке и транспортировке горной массы;
- разработка, изготовление и внедрение полевой и стационарной аппаратуры для контроля и управления качеством минерального сырья в рудах, пульпе и концентрате;
- предоставление сервисных услуг по гарантийному обслуживанию, ремонту и модернизации геофизического оборудования;
- предоставление услуг по каротажу шарошечных и колонковых скважинах по определению содержания магнитного и общего железа. [2]

Все системы и устройства активно функционируют во многих горно-металлургических комбинатах, заводах, карьерах, шахтах и т.д. Их эффективность и польза проверенна годами и подтверждена потребителем. Однако, время не стоит на месте и возникает новая проблема – автоматизация передачи результатов исследований с рабочих станций на сервер, возможность просмотра информации об измерениях и использования данных удаленным пользователем в режиме реального времени для управления горными работами, ведения контроля качества и оформление отчетов.

Актуальность проблемы контроля качества сырья

Добыча и обогащение руды является трудоемким и сложным процессом. В настоящее время на каждом этапе происходит перераспределение качества минерального сырья в рудах при помощи специализированных устройств и методов. Это вызвано большим количеством факторов влияющих на потерю качества во время процесса добычи и переработки. Одним из них является «человеческий фактор». В настоящее время все результаты исследований снимаются вручную и в последующем передаются далее, а значит и возможность ошибки не исключена. Для решения этой проблемы необходимо произвести автоматизацию процесса снятия показаний с приборов и отправки их на сервер. Так же необходимо обеспечить возможность удаленного контроля и управления горными работами, что значительно облегчит и ускорит процесс производства, улучшит качество конечного продукта и сырья на промежуточных этапах. В отличии от своих аналогов система обеспечивает ведение контроля качества железорудного сырья на всех этапах добычи и переработки, отслеживает результаты исследований в режиме реального времени и за определенный период, обеспечивает возможность составления отчетов в виде текста, графиков и таблиц. Данная задача является глобальной и востребованной. Чтобы решить ее, было принято разработать проект, а в последующем разработать рабочую автоматическую систему по контролю и решения управлению качеством железорудного сырья на всех стадиях технологической цепи.

Постановка задачи

Учитывая актуальность проблемы и возможности реализации решения, была сформирована основная цель проекта: создание комплексной системы оперативного контроля качества железорудного сырья, объединяющую все локальные системы на различных этапах добычи и переработки; разработка автоматизированной системы управления и обработки данных, полученных от устройств по контролю качества сырья, находящиеся в эксплуатации на производстве; передача информации между узлами системы; шифрование и хранение результатов исследований в базе данных (БД); разработка программного обеспечения (ПО) для взаимодействия пользователя и сервера, ведение статистики и контроля качества по добыче железорудного сырья, формирование отчетов.

Функциональные узлы

Разрабатываемая автоматизированная система представляет собой сложную систему взаимодействия основных функциональных узлов с использованием различных способов связи на разных этапах работы. Основными функциональными узлами являются:

1. Станция является важным узлом системы. Ее основной задачей является получение значения качества железорудного сырья и отправка данных исследования удаленному серверу.

В качестве станций предполагается использовать:

− каротажная станция «Карьер-Кривбасс» (многофункциональная мобильная станция предназначена для одновременного оперативного контроля (каротажа) содержания железа магнитного и общего в шарошечных и колонковых скважинах глубиной до 30 м);

− ПАКС-5КК (устройство предназначено для каротажа скважин диаметром до 300 мм с целью определения распределения содержания железа магнитного по глубине скважины. Критерием содержания железа является величина магнитной восприимчивости пород);

− ПАКС-5МК (устройство предназначено для экспрессного определения железа магнитного в навалах дробленной руды, обеспечивает определение содержания магнетита, предоставляет возможность

длительного хранения средних значений содержания полезного компонента);

− ДЖМ-4 (устройство предназначено для экспрессного определения массовой доли железа магнитного в пробах дробленого рудного материала).[3]

Каждая станция должна состоять из таких основных элементов:

− датчик железа;
− память устройства;
− модуль связи;
− внешний интерфейс устройства.

В целях проекта необходимо разработать принципиальную схему и печатную плату модуля передачи данных. Так как в условиях производства использование оптоволоконной сети не допустимо, а передача информации при помощи Wi-Fi крайне затруднительна, в связи с малым радиусом действия и слабой проходимостью сигнала, было принято решение использовать GSM канал связи. Его преимущество - это независимость расстояния между объектами связи, а так же возможность доступа в всемирную систему Internet и использование TCP/IP протоколов связи, что облегчает разработку ПО системы, обеспечивает надежность и быстродействие передачи данных. В связи с этим в качестве модуля связи станции должен выступать встраиваемый GSM-модуль и GSM-интерфейс для полного его функционирования. Важным и обязательным фактором работы системы является наличие GSM-покрытия сотовой сети оператора связи в зоне работы устройств. Можно использовать мобильную базовую станцию в качестве соединительного узла между устройством и сервером, однако это очень дорогостоящая система приборов, которая не окупает затрат на ее приобретение. Как следствие, в случае отсутствия GSM-покрытия оператором сотовой связи отправка данных на сервер будет не возможна.

Каждая рабочая станция должна обеспечивать следующие возможности:

− сбор данных о результатах исследования качества сырья;
− получение координат местоположения при помощи GPS;
− кодирование и шифрование данных;
− формирование к отправке по защищенному GSM каналу;
− интерактивный обмен пакетами с удаленным сервером системы.

Так же при разработке должна быть учтена помехозащищенность канала связи в условиях промышленного производства и защита от внешнего проникновения и перехвата пакетов.

2. Сервер должен быть реализован как удаленная ЭВМ с возможностью доступа к глобальной сети Internet. Сервер является ключевым узлом всей системы, т.к. предоставляет возможность связи между клиентом, станциями и администратором. Так же на сервере находиться БД со всеми результатами исследования, полученными на протяжении работы системы. На машине сервера должна быть установлено разработанное ПО для взаимодействия всех узлов системы.

Основные функции сервера:

− прием и передача данных между узлами;
− хранение данных исследований в БД;
− хранение данных пользователей в БД;
− идентификация пользователей и рабочих станций;
− интерактивность со всеми узлами системы;
− обработка запросов пользователей и администратора системы;
− кодирование и шифрование данных;
− проверка и подтверждение целостности приема и передачи пакетов данных;
− ведение истории запросов пользователей и добавления данных исследований.

Связь между узлами системы происходит при помощи TCP/IP протокола, что позволяет сделать обмен пакетами данных быстрее, проще и надежнее.

Важным фактором для функционирования системы является наличие статического внешнего IP-адреса. Возможность подключения следует уточнить у Internet-провайдера, предоставляющего услуги Internet-связи предприятию. В случае его отсутствия реализация системы становиться крайне затруднительной.

Серверное ПО должно быть выполнено как консольное или Windows Forms приложение. Для большей наглядности и удобства рекомендуется использовать второй тип, т.к. это упрощает использование ПО, делает его более гибким, интуитивно

понятным и привычным обыденному пользователю ПК.

Программное обеспечение будет выполнено на платформе Net Framework 4.x и языке программирования C#. Выбор обоснован его новизной, возможностями, быстродействием и постоянным развитием. На данный момент необходимо использовать операционную систему (ОС) семейства Windows. При использовании ОС сторонних производителей система не будет функционировать.

Для обеспечения защиты информации доступ к машине сервера должен быть ограничен доверенным кругом лиц. Для обеспечения защиты внутри сети будет использоваться система идентификации пользователей.

3. Пользователь (Клиент) представляет собой управляемый человеком ЭВМ с установленным специализированным ПО для взаимодействия и интерактивности с удаленным сервером системы. Клиентом является любой пользователь сети, который авторизировался в системе при помощи ПО. Принцип работы заключается в формировании запроса пользователем, отправка его на сервер и ожидание результатов выборки.

Клиентское ПО должно предоставлять следующие минимальные возможности:

- доступ к глобальной и/или локальной сети Internet;
- доступ к портам сети необходимым для взаимодействия сервера и клиента;
- возможность ведения интерактивности с сервером системы;

4. Администратор. Узел администратора представляет собой управляемую человеком удаленную ЭВМ с установленным специализированным ПО и возможностью доступа ко всем функциональным узлам системы. Администратором системы может быть как доверенное лицо предприятия, так и лицо со стороны разработчиков.

Администратор необходим для контроля и удаленного обслуживания системы в случае возникновения проблем при работе клиентского или серверного узла. В случае необходимости, администратор при помощи разработанного ПО может:

- получить доступ к БД хранящимся на сервере;
- отправка и получение пакетов от сервера;
- возможность регистрации и авторизации пользователей системы;
- выбор критериев поиска для формирования запроса серверу;
- формирование результатов поиска в табличной и графичной форме представления;
- возможность сохранения результатов запроса в электронной и печатной форме;
- возможность составления и формирования отчетов;
- возможности ведения статистических данных, формирования графиков и таблиц для анализа и контроля качества сырья на всех этапах производства.

Так же, важным фактором является наличие на каждой машине клиента Net Framework 4.x. В случае отсутствия такового его необходимо установить.

Связь между клиентом и сервером должна происходить посредством TCP/IP протокола. Между каждым узлом системы «клиент-сервер» необходимо постоянное, качественное соединения, дабы избежать потерю и повреждение данных во время обмена пакетами.

Стоит отметить, что подключение к серверу возможно как внутри сети предприятия, так и извне. Это позволяет технологу, независимо от его текущего местоположения, при помощи глобальной сети Internet контролировать процесс добычи и обогащения на всех этапах. Это значительно упрощает процесс контроля, улучшает организацию труда и позволяет моментально реагировать на поступающие данные.

- просматривать историю запросов пользователей;
- просматривать историю добавления данных исследований;
- выполнить необходимые операции по выборке из БД;
- включать/выключать/перезапускать ПО сервера;
- просмотреть список подключенных пользователей;
- отключать при необходимости пользователей;
- рассылать сообщения как конкретному пользователю, так и всем одновременно, находящимся в сети.

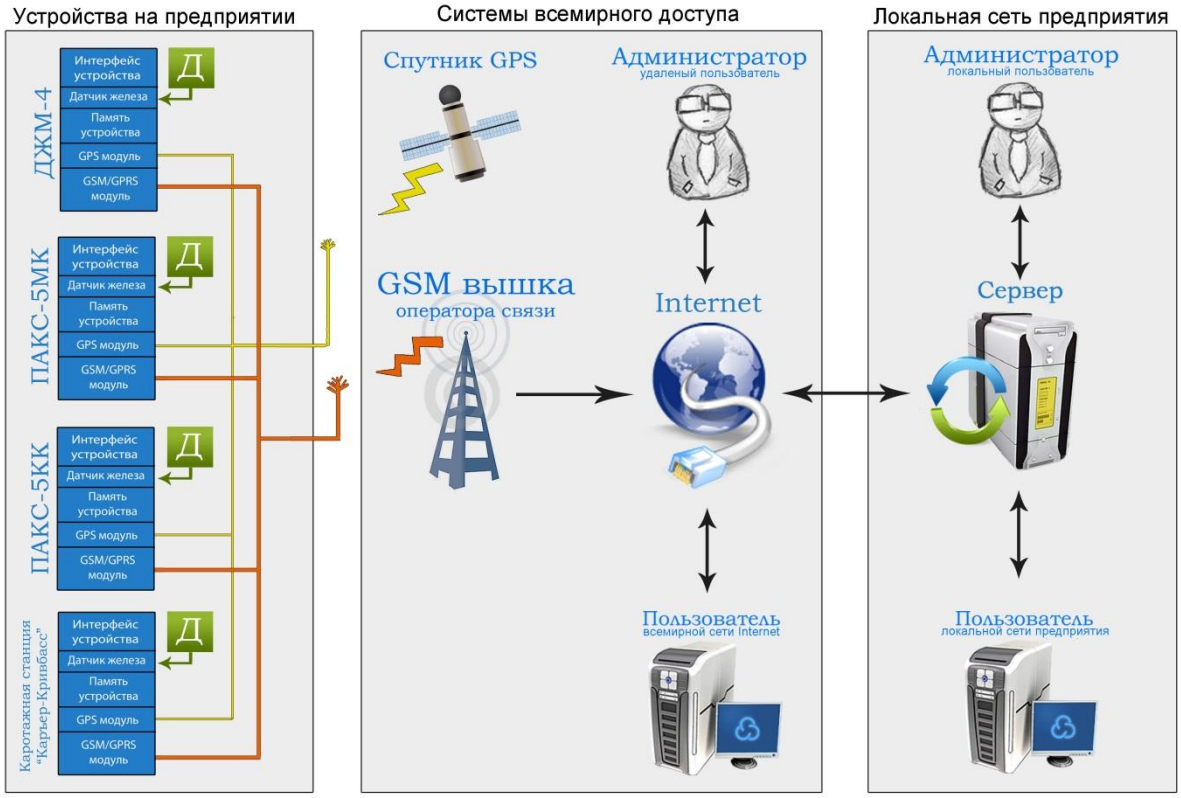

Рис. 1. Функциональная схема работы автоматизированной системы управления горными работами и контроля качества железорудного сырья

Функциональная схема

Рассмотрев и определившись с основными функциональными узлами была разработана функциональная схема автоматизированной системы управления горными работами и контроля качества железорудного сырья (рис. 1).

На первом этапе работы получаем, при помощи специализированных устройств и оборудования, содержание полезного компонента в процентном эквиваленте. Как только операция закончена, устройство автоматически определяет координаты своего местоположения со спутника GPS и отправляет данные на сервер. На отрезке «устройство-сервер» информация передается по GSM-каналу связи - определяется ближайшая базовая станция оператора сотовой связи; результаты исследования передаются по GSM-каналу на нее, а после перенаправляются непосредственно через всемирную сеть Internet. При поступлении на сервер, результаты проверяются и сохраняются для последующей возможности работы с ними. При необходимости любой зарегистрированный пользователь системы может запросить необходимую информацию находящеюся на сервере и получить результаты в виде таблиц и графиков, как за конкретный промежуток, так и в режиме реального времени. Пользователь может находиться внутри сети предприятия и во всемирной сети Internet. Администратор системы имеет более расширенные возможности доступа и управления, чем пользователь. Вся система тщательно защищается от внешнего проникновения.

Алгоритмы работы системы

В этой статье рассмотрены два основных этапа работы системы по контролю качества железорудного сырья:

- передача и получение информации с устройства на сервер;
- получение, обработка, и отправка результатов запроса пользователя.

При рассмотрении будет представлена блок-схема каждого алгоритма с детальным пошаговым описанием. Алгоритмы работы с

БД и получение значения качества железорудного сырья рассматриваться не будут, т.к. это не является основной задачей статьи.

1. Алгоритм передачи и получения информации с устройства на сервер.

Передача информации между устройством и сервером происходит посредством двух типов связи, а именно с помощью GSM-канала и Internet. Каждое устройство системы имеет свой алгоритм определения количества полезного компонента в железорудном сырье. В данном разделе мы рассмотрим общую блок-схему алгоритма получения данных и отправка их на сервер (рис.2).

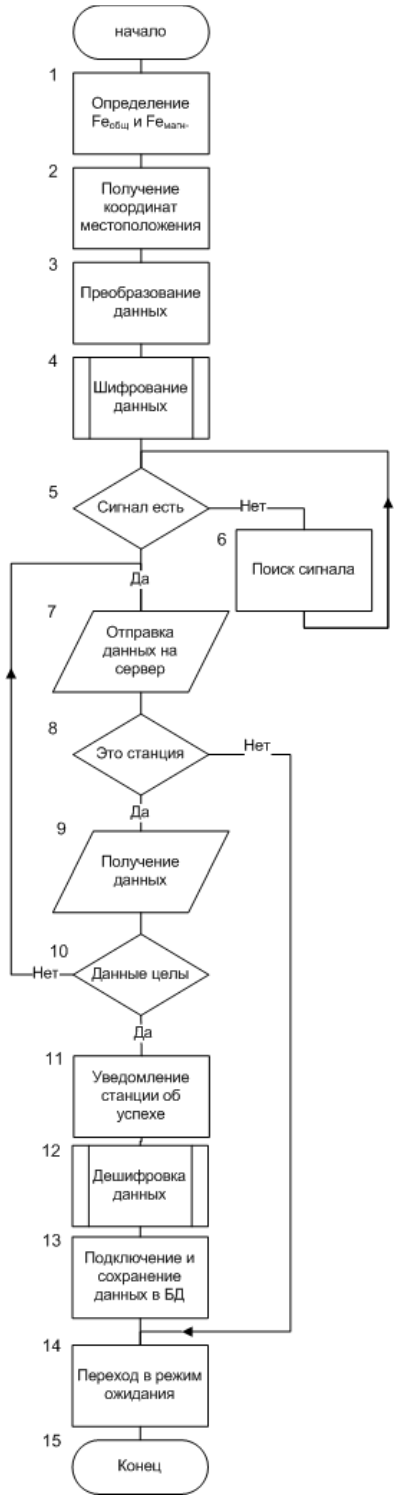

Рис. 2. Блок-схема алгоритма передачи и получения информации от устройства к серверу

Первым этапом работы станции является определение и вычисление усредненных значений $Fe_{общ}$ и/или $Fe_{магн}$ (1). После того как результаты исследования получены, станция определяет свое местоположение использую технологию GPS (2). Все данные необходимо подготовить к отправке, поэтому происходит их преобразование до нужной формы (3). В целях безопасности информацию необходимо шифровать согласно одному из распространенных методов кодирования(4). Так же следует проверить возможность подключения к серверу и уровень сигнала, в случае проблем необходимо устранить проблему, а именно проверить антенну станции и немного изменить расположение и найти точку, где сигнал присутствует (если это возможно)(5, 6). По окончанию формирования данных и проверки подключения информация передается на удаленный сервер, после чего станция переходит в режим ожидания ответа об успехе передачи (7). При запросе на прием от станции, сервер идентифицирует станцию и, в случае успеха, получает результаты исследования (8, 9). По окончанию приема происходит проверка данных на целостность (10). Если данные повреждены, то сервер уведомляет об этом станцию и ожидает повторной передачи. Если же данные целы и не модифицированы, то сервер дешифрует их и уведомляет станцию об успехе - станция вновь готова к отправке (11, 12). На следующем этапе сервер подключается к БД, присваивает данным уникальный идентификатор и сохраняет их (13) По завершению сервер снова становиться в режим ожидания запросов (14). Так же предусматривается хранения информации в локальной БД станции на случай не корректного завершения работы устройства и просмотра ранее проведенных исследований.

2. Алгоритм приема, обработки, и передачи результатов запроса пользователя.

Важным этапом и основной возможностью системы является участок «пользователь-сервер». На этом промежутке цепи предоставляется возможность каждому удаленному пользователю произвести не явную выборку по необходимым параметрам для последующего анализа, ведения оперативного контроля качества и формирования отчетных данных. Стоит заметить, передача информации происходит по TCP/IP протоколу, что значительно облегчает организацию пересылки пакетов данных для разработчиков. Блок-схема алгоритма и пошаговое описание представлено ниже (рис.3).

При запуске ПО пользователю необходимо авторизироваться в системе с помощью ранее созданного администратором сети логина и пароля. Для каждого пользователя присваивается уникальные параметры входа (1). После того как авторизация прошла успешно, клиент может выбрать тип объекта и параметры поиска (время, период, показатели и т.п.)(2). Выбрав нужные параметры, запрос формируется и отправляется по защищенному каналу на сервер (3). Во время поиска сервером данных, клиент переходит в режим ожидания. При получении запроса на выборку от клиента сервер подключается к БД и производит поиск(4, 5). В случае если запрашиваемая информация отсутствует, сервер уведомляет об этом клиента и переходит в режим ожидания для последующего приема и обработки запросов (6, 7). Если же запрашиваемая информация найдена в БД, сервер шифрует ее (8), преобразовывает для отправки (9) и, по окончанию всех этапов, отправляет (10). После чего ожидает возвращение флага успешности операции. Если результат положителен, то сервер возвращает в режим ожидания запросов от клиентов. При получении результатов от сервера (11), клиентское ПО проверяет целостность данных (12) и в случае их повреждения серверу отправляется сообщение о проблеме и повторный запрос на получение. Если данные прошли проверку целостности, они дешифруются (13) и отправляется сообщение серверу об успехе операции (14). После того как данные полностью дешифрованы, они представляются пользователю в виде таблицы с последующей возможность работы с ними (формирование отчетов, ведение статистики, построение графиков для анализа) (15). По окончанию пользователь может сделать новый запрос на выборку (2-16).

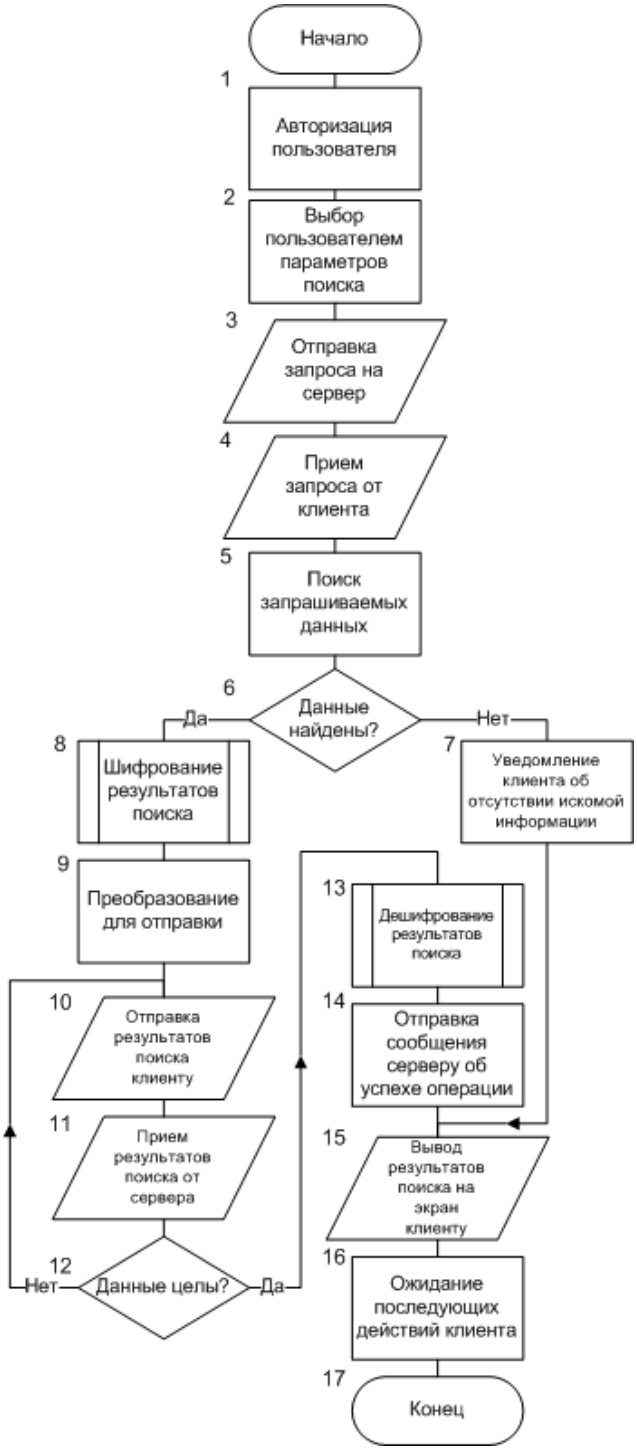

Рис. 3. Блок-схема алгоритма получения, обработки и отправки результатов запроса пользователя

Работа с каждым пользователем будет происходить в отдельном потоке, что позволит распараллелить работу сервера и значительно увеличить быстродействие системы. При поступлении запросов от пользователей и станций, на сервер будет сохраняться история в которую будет заноситься дата, время, идентификатор пользователя/станции, IP-адрес, строка запроса для улучшения контроля и управления работами администратором сети.

Выводы

В статье был проведен обзор проекта по автоматизации оперативного контроля качества железорудного сырья и управления горными работами. Была определена актуальность проблемы во время добычи и переработки минерального сырья. Кратко рассмотрена функциональная схема системы и минимальные требования к ее ключевым узлам. Так же были представлены основные алгоритмы работы системы с подробным пошаговым описанием и графическим представлением в виде блок-схемы. Информация в статье является кратким обзором основных положений для реализации проекта автоматизации оперативного контроля железорудного сырья на всех этапах добычи и переработки разрабатываемой ООО «РУДПРОМГЕОФИЗИКА» совместно с проблемной лабораторией Криворожского национального университета и ООО «НПП «Тетра».

Список литературы

1. А.А.Азарян Состояние проблемы контроля качества руд при добыче и переработке железорудного сырья. / А.А.Азарян, Ю.Г Вилкул, В.А Колосов и др. // Металлургическая и горнорудная промышленность №1, Днепропетровск. 2004.-С 88-90.

2. Спосіб градуювання зондів гамма-гамма каротажу свердловин які частково заповнені водою: патент №39985 - Україна / А.А.Азарян, В.Є.Василенко, О.В.Швидкий; власник патенту ТОВ «Рудпромгеофізика».

Пристрій для оперативного контролю масової частки заліза магнітного у гірській масі : патент 80694 / Азарян А.А., Дрига В.В., Цибулевський Ю.Є. ; власник патенту ТОВ «Рудпромгеофізика».

Сведения об авторах

Азарян Альберт Арамаисович, доктор техничных наук, профессор, заведующий кафедры моделирования и программного обеспечения, Криворожский национальный университет, директор ООО «Рудпромгеофизика»; научные интересы: информационное обеспечение геофизических методов оперативного контроля и управления качеством минерального сырья.

e-mail: aza1207@ya.ru

Азарян Владимир Альбертович, кандидат технических наук, доцент, кафедра открытых горных работ, Криворожский национальный университет, заместитель директора ООО «Рудпромгеофизика»; научные интересы: оперативный контроль и управление качеством минерального сырья.

e-mail: astp165@ukr.net

Гриценко Андрей Николаевич, аспирант, кафедра моделирования и программного обеспечения, Криворожский национальный университет; научные интересы: оперативный контроль и управление качеством минерального сырья.

e-mail: andryuha_person@mail.ru

Кайгородов Руслан Александрович, инженер-программист, ООО «Рудпромгеофизика»; научные интересы: информационные технологии, разработка программного обеспечения для оперативного контроля и управления качеством минерального сырья.

e-mail: kissblog@mail.ru

Мирошник Денис Юрьевич, инженер-программист, ООО «Рудпромгеофизика»; научные интересы: информационные технологии, разработка программного обеспечения для оперативного контроля и управления качеством минерального сырья.

e-mail: daldenis@yandex.ru

УДК 004.04

Лисак Володимир Васильович,
Київської державної академії водного транспорту ім. гетьмана П. Конашевича – Сагайдачного

ІНЖЕНЕРІЯ АНАЛІТИЧНИХ ЗНАНЬ

Розглянуто еволюцію аналітичних систем.Запропоновано поняття аналітичних знань як результатів методів обробки даних, а такожінженерний підхід до їх отримання з баз даних завдяки представленню всіхметодів обробки у виглядіграфічних об'єктів. Об'єкти мають клієнт-серверну архітектуру, інкапсулюють в собі вхідні дані, методи обробки даних з обраними режимами обробки, а також результати у вигляді таблиць, графіків та схем. Серед об'єктів представленіприклади методів підготовки даних (вибірки з різних джерел, узгодження, перевірка на коректність, вирізування даних з таблиць, транспонування тощо), а також методів пошуку знань та залежностей. Будучі одночасно і клієнтом і сервером аналітичні знання можуть створювати ланцюжки виконання методів. Послідовне застосування методів створило можливість концентрувати аналітичну інформацію і отримувати комбіновані аналітичні знання.

Рассмотрена эволюция аналитических систем. Предложено понятие аналитических знаний как результатов методов обработки данных, а также инженерный подход их получения из баз данных благодаря представлению всех методов обработки в виде графических объектов. Объекты имеют клиент-серверную архитектуру, инкапсулируют в себе входные данные, методы обработки данных с выбранными режимами обработки, а также результаты в виде таблиц, графиков и схем. Среди объектов представлены примеры методов подготовки данных (выборки из различных источников, согласования, проверка на корректность, выборки данных из таблиц, транспонирование и т.д.), а также методов поиска знаний и зависимостей. Будучи одновременно и клиентом и сервером аналитические знания могут создавать цепочки выполнения методов. Последовательное применение методов создало возможность концентрировать аналитическую информацию и получать комбинированные аналитические знания.

The evolution ofanalytical systemshas been considered.The concept of analytical knowledge as a result of data processing and engineering approach to their receipt of databases through presentation of all methods of treatment in the form of graphic object. Objects have a client-server architecture that encapsulate a data input, data processing with selected modes of processing, and the results in the form of tables, graphs and diagrams. Among the objects to be methods of training data (samples from different sources, alignment, checking for correctness, cutting data tables, transpose, etc.), as well as methods of seeking knowledge and relationships. Being both a client and server analytical knowledge can create chains of execution methods. Consistent application of methods created to concentrate analytical information and receive a combined analytical knowledge.

Ключові слова:аналітичні системи, інженерія знань, методи пошуку знань

Вступ

Сучасне суспільство все більше називають суспільством, що базується на знаннях, тим самим підкреслюючи важливу роль знань, науки та інформації в суспільному житті. Для сьогодення характерна глобалізація змін в економіці, науці, освіті завдяки поширенню нових технологій здійснення та пришвидчення комунікацій. Сьогодні вже процеси виробництва та інженерії знань не виступають окремо, а стають необхідною фундаментальною умовою функціонування та розвитку суспільства.

В той же час сучасний діловий світ розуміє, що без глибокого аналізу інформації, яка переповнює ринки та внутрішню діяльність організацій, неможливе успішне ведення бізнесу. Потоки інформації при її кваліфікованій обробці, аналізі та синтезу висновків, здатні надати підприємству конкурентні переваги по відношенню до інших учасників ринку, які її не мають.

Актуальність розвитку аналітичних комп'ютерних систем підтверджується тенденціями напрямків розробок у світових лідерів програмного забезпечення, а саме: на ринки програмного забезпечення виходять інформаційно-аналітичні системи для використання у різних сферах. Аналітичні системи – BI (Business Intelligence[1]) – представлені цілою лінійкою рішень більш ніж 10 різних компаній. За думкою експертів IDC[2],

[1]Business intelligenceабо скороченоBI — бізнес-аналіз, бізнес-аналітика. Під цим поняттям частіше всього розуміють програмнезабезпечення, призначене для допомоги менеджеру в аналізіінформаціїпро свою компанію.

еволюцію інформаційно-аналітичних систем можна розділити на 3 хвилі:

1 – (до 1990 року) збір інформації та підготовка звітності;

2 – (1990-2005 рік) розвиток швидкого багатомірного аналізу на базі технології OLAP [1], а також самостійне створення нерегламентованої звітності;

3 – (з 2005 року) створюється акцент на розвиток прикладних засобів застосування, що включає аналіз, прогностику та пошук прихованої інформації (методи Data Mining [2]).

Сьогодні мияк раз бачимо появу BI-решень третьоїхвилі. До них можна, наприклад, віднести BI-системивідвідомихсвітових лідерів: Oracle, IBM, Microsoft, SAS тощо.

Що стосується перспектив використання технологій, то зараз пост-радянський ринок як раз стоїть на порозі початку впровадження BI-систем третьої хвилі, які здатні шукати приховану інформацію, будувати прогнозну аналітику та проводити перехресний аналіз інформації з несумісних на перший погляд джерел даних. Самі експерти IDC вважають, что ці хвилі мають 15-річний цикл, але, приймаючи до уваги швидкість, з якою на світовому ринку з'являються нові решення, можна очікувати, что третя хвиля все ж закінчиться раніше і вже через пару років пост-радянський ринок стане свідком появи рішень нового, четвертого покоління. Можна очікувати, що це буде більш тісна інтеграція систем класу підтримки прийняття рішень DSS (Decision Support System), систем прогнозування та пошуку прихованих залежностей (Data Mining). Іншими словами, користувачу буде надана можливість вибору сценарія розвитку ситуації, виходячи з якого система сама проведе аналіз накопиченої інформації, побудує прогноз зміни ключових показників та запропонує оптимальні варіанти дій, які б привели до ліпшого результату. Тобто BI-система позбавить користувача від необхідності виконання довгої рутинної роботи з пошуку причинно-наслідкових зв'язків при аналізі даних, передачі результатів роботи однієї системи в іншу, контролі коректності завантаженої інформації тощо. Ці завдання будуть виконуватися автоматично - від користувача буде потрібно всього лише на початку роботи вибрати сценарій розвитку тієї чи іншої задачі, а в кінці - найбільш вподобаний йому оптимальний варіант вирішення цього завдання, що відповідає змісту фрази "Business Intelligence".

Інформаційно-аналітичні системи являють собою надбудову над тими інформаційними системами, що вже функціонують на підприємствах, вони не впливають на їх функціонування та не вимагають їх заміни. Ключовою функцією цих систем є обробка, акумулювання та аналіз даних з усіх видів діяльності організації чи окремої системи, отримання закономірностей, залежностей, прогнозування та оцінці можливих ризиків, оптимізації діяльності.

Фактично функціями таких систем є з одного боку - робота з даними (пошук, формування вибірок з таблиць, їх склеювання, розрізування, транспонування, створення зведених та агрегованих таблиць), а з іншого - їх перетворення до вигляду, з якого можна отримати нові знання в тематичній області після застосування різних математичних методів або практичних алгоритмів, які застосовуються для прийняття рішень або системного моделювання процесів чи явищ.

Метою цієї статті є опис та реалізація інженерного підходу до створення аналітичних знань, відокремлення інженера знань від експерта у тематичній області та математика з його поглибленими знаннями у математичних методах обробки даних. Таким чином робота з даними для отримання аналітичної інформації (знань) повинна стандартизуватись та формуватись як інженерна дисципліна з основним виконавцем – інженером знань.

Уточнення поняття «аналітичні знання»

З розвитком структур виникла концепція *знань*, у відповідності до якої об'єднуються блоки різних типів інформації, а перехід від даних до знань - логичний наслідок розвитку та вдосконалення інформаційно - логічних структур, що оброблюються на ЕОМ [3]. Що ж таке «знання»? Поняття «знання» різними авторами трактується по-різному, але як таке,

[2] IDC - International Data Corporation є провідним світовим постачальником ринкової інформації, консультаційних послуг та організатор заходів на ринках інформаційних технологій, телекоммунікацій та користувацької техніки. Більш ніж 1000 аналітиків IDC, глобальний, регіональний та місцевий досвід в области технології в більш ніж 110 країнах світу.

що найбільше відповідає тематиці статті, то це - виявлені закономірності в предметній області (принципи, зв'язки, закони), що дозволяють розв'язувати тематичні задачі. Поняття «знання» прийнято поділяти на декларативні та процедурні. Декларативні знання - це певна множина тверджень, незалежно від того де і коли вона використовується. **Процедурні знання** або **правила** являють собою набір певних процедур перетворення даних в знання. Поділ знань на декларативні та процедурні достатньо умовний і відомі моделі представлення знань використовують у різній мірі ті або інші поняття.

Формування знань виконується із залученням експерта – фахівця в тематичній області. А задача інженера - це виявити каркас поглядів та висновків експерта. Інженер із знань може спиратися на дві найбільш популярні теорії мислення – логічну та асоціативну. Якщо логічна, як правило, використовується в розробках з штучного інтелекту, то асоціативна будується з минулого досвіду, методом спроб та помилок, а також методом аналогій. І тут можна говорити про аналітичні *знання,* які принципово відрізняються від інших перш за все способом їх отримання. Як назвати результати застосування однакових процедурних знань до різних даних? Звичайно отримані результати будуть різні, і вони будуть претендувати на назву *аналітичні знання*. Будемо вважати, що аналітичні знання - це знання, які отримані як результат застосування процедурних знань, а саме: методів DataMining[3] (розв'язок задачі пошуку асоціацій, послідовностей, задачі класифікації, кластерізації, задач прогнозування та оптимізації) або методів багатомірної статистики, що застосовуються до статистичних даних в базах даних, або певних моделей, наприклад, нейромережевих моделей чи генетичних алгоритмів тощо.

[3]Однез найважливішихпризначеньметодівData Miningполягає внаочному поданнірезультатів обчислень, щодозволяє використовуватиінструментарійData Miningлюдьми, які не мають спеціальноїматематичної підготовки.

Системи аналізу та обробки наукової інформації, або системи отримання аналітичних знань будуються на принципах та технологіях, відмінних від принципів побудови інших інформаційних систем сучасного підприємства, тому що дані, які використовуються, вимагають проведення спеціальної їх підготовки, надання їм коректності для подальшого використання.

Постановка проблеми.

Користувачі всередині організацій потребують різних типів інтерфейсів в залежності від типу інформації, з якою вони працюють. Керівники потребують спеціальної інформації, що подається за допомогою інтуїтивно зрозумілого інтерфейсу, з точними показниками досліджуваних експериментальних даних. Бізнес-аналітикам необхідна можливість працювати безпосередньо з даними, використовуючи інтерфейс орієнтований на певну задачу.

Необхідно створити таке універсальне середовище - базу *аналітичних знань*, в якій можна виконувати елементарну обробку даних: використовувати вибірки даних для аналізу з різних, можливо зовнішніх, сховищ, узгоджувати їх для подальшої обробки, виявляти та автоматично коригувати їх структуру, застосовувати методи отримання знань та зберігати у вигляді аналітичних знань. І все це завдяки графічному інтуїтивно зрозумілому інтерфейсу, який буде зручним як для керівника організації, бізнес-аналітика, так і для інженера знань.

Необхідно також щоб аналітична система не вимагала спеціальних математичних знань, але дозволяла галузевому аналітику або інженеру знань виконувати пошук знань та залежностей за допомогою новітніх методів.Окрім того для вдосконалення отримання аналітичних знань необхідно, щоб без втручання «програміста» можна було б в одному результаті послідовно акумулювати результати послідовної роботи окремих методів обробки.

Основні принципи побудови аналітичної системи

Для того, щоб система управління данними претендувала на назву аналітичної, необхідно її спроектувати як базу «знань» з різними методами обробки даних та пошуку знань.

Отже, якою повинна бути базова одиниця інформації в базі аналітичних знань? Легше за все можна вивчити роботу аналітиків (інженерів знань), де виконується кропітка робота з підготовкою даних (вибірки з різних джерел, узгодження, перевірка на коректність, вирізування даних з таблиць, транспонування тощо) – всі ці методи обробки повинні бути присутні в аналітичній базі знань і використовуватись як універсальні аналітичні об'єкти на рівні з процедурними знаннями. В той же час, якщо взяти будь-який аналітичний звіт – в ньому присутні результати математичних експериментів, статистичних методів оцінки окремих змінних, їх порівнянь, знаходження залежністей, групування або класифікація за різними ознаками, які зменшують кількість змінних створюючи більш складні структури. Застосовуються також різні моделі, методи прогнозування та оптимізації.

Таким чином основним елементом такої бази знань повинен бути ***графічний об'єкт***, який об`єднує метод обробки чи алгоритм розрахунку з можливістю динамічного підключення вхідних даних з різних джерел та збереження результатів застосування методу обробки як у вигляді таблиці, так і у вигляді схеми або графіка, в зрозумілій, легко прийнятній аналітиком формі(дивись рис.1). Такий об'єкт будемо називати «аналітичним знанням» або просто «знанням».

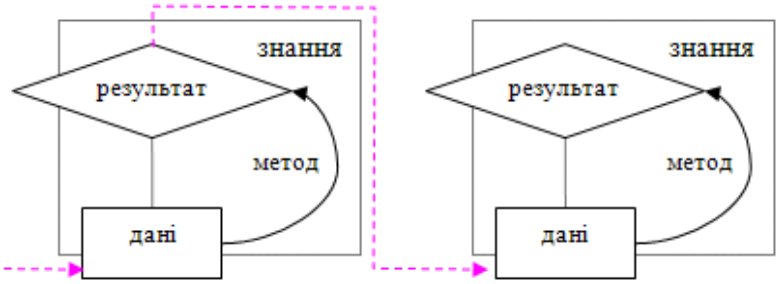

Рис.1.Схема базових елементів «знання», які включені в послідовний ланцюжок розрахунків

Складовими елементами «знання» є:

- дані – вхідна інформація, яка необхідна для виконання певного методу розрахунку;
- метод (розрахунку) – деякий статистичний, аналітичний або математичний метод, який є серед множини існуючих, а також той, який завжди можна додати до використання; це може бути також метод обробки та підготовки даних;
- результат – інформація або дані, які отримані шляхом застосування методу обробки чи розрахунку до вхідних даних. Результат може слугувати вхідною інформацією для іншого знання, забезпечуючи послідовну обробку різних методів.

Кожний об'єкт - «знання» забезпечено стартером для автоматизованого запуску на виконання поточного об'єкту в залежності від сигналу «виконано» попередніх об'єктів. Таким чином забезпечується потокова обробка «знань», яка формується автоматично відповідно до створення сценарію послідовної обробки даних інженером знань (дивись рис.2)

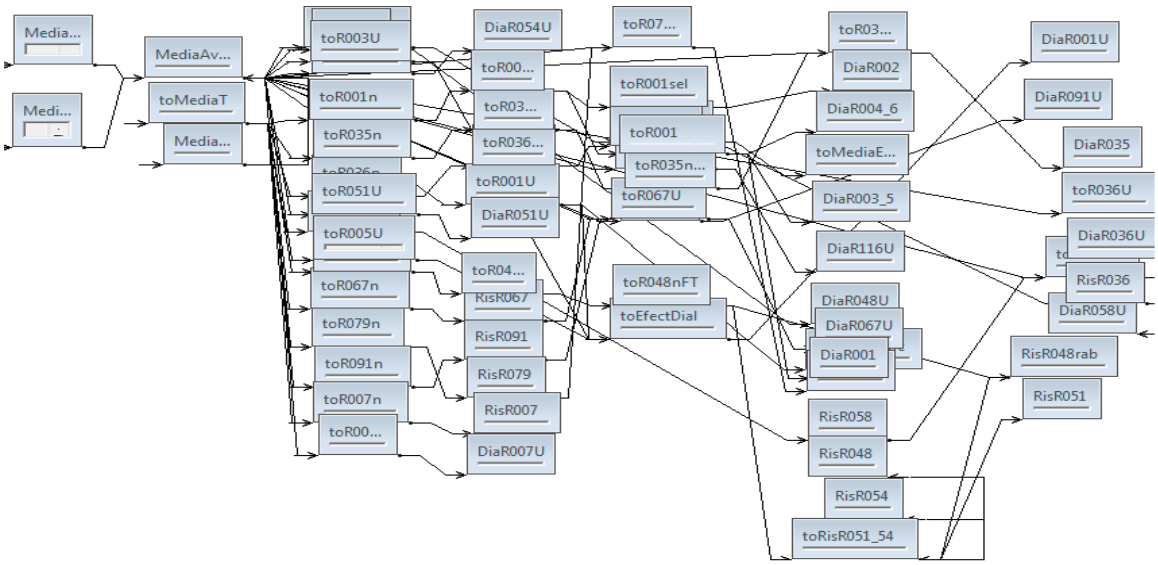

Рис.2.Схема потокової обробки «знань», яка формується автоматично

Розвиток базового об'єкту в майбутньому не торкнеться його суті, але можливо його розширити і наділити декількома результатами, а також поєднати декілька методів послідовно для утворення нового методу - ланцюжка методів.

Інженерія аналітичних знань повинна включати стандартні підходи до обробки даних, зокрема, якщо виконується аналіз невідомих данних.

Системний підхід до аналізу даних вимагає, щоб з одного боку об'єктами в базі були елементарні дії роботи над перетворенням таблиць, а з другого - методи пошуку знань.

Засоби попередньої обробки даних, які для інженера знань забезпечуть всі можливі операції з даними, і є відносними, це:

- запити на вибірку із зовнішніх таблиць, об'єднання таблиць, склеювання, транспонування, створення зведених таблиць, чи навпаки розведених (з прямокутних таблиць створюються подовжені), при чому дії без явного вказування назв полів;

Після підготовки даних до них можна застосовувати системний аналіз даних [4], різноманітні методи аналізу та перетворення даних в знання, саме:

- дослідження структур даних (основні статистики, групові середні, частотний розподіл, таблиці перерізів), їх порівняння;
- виявлення статистично-значущих зв'язків між змінними, методи асоціацій та послідовностей (кореляційний аналіз, факторний аналіз, парна регресія);
- групування, класифікація та кластерізація (кластерний аналіз, дерево рішень, нейромережі);
- знаходження факторів для зменшення змінних (факторний аналіз);
- визначається поведінка окремих груп, прогнозування змін, порівняльний аналіз (екстраполяція, динамічні ряди, нейромережі, множинна регресія);
- методи оптимізації (метод Тагучі та нейромережі);

Якщо порівнювати інженера знань, який створює ланцюжки обробки даних та отримання знань з осмисленою роботою науковця, то необхідно аналітичну базу знань забезпечити можливістю оформлення в один об'єкт ланцюжка методів, що повторюється. Таким чином в системі підтримується концепція «виникнення» нових складних структур обробки даних. Це також вписується в поняття того, що будь-який розвиток припускає споживання найпростіших елементів для поступової трансформації і переходу на більш високий рівень[5].

Результати.

Таким чином було розроблене ядро системи з потоковою обробкою даних, в якій база аналітичних знань являє собою набір засобів доступу до зовнішніх стандартних баз даних, забезпечує створення вибірок, засобів обробки даних, методів аналітичної обробки. Одиницею такої бази є складний об'єкт, який інкапсулює в собі вхідні дані, методи обробки та їх режими, а також результат обробки. При

цьому результат представляється не тільки у вигляді розрахунку, а і у візуальній формі – у вигляді графіка, діаграми чи схеми. В ній поряд із поширеними стандартними методами обробки даних, методів математичної статистики та пошуку знань, використовуються розроблені нові аналітичні моделі, як результат послідовного розрахунку декількох методів. Система має графічний інтерфейс, зручний для користувачів різного професійного рівня.

Приклад найпростішої потокової обробки даних, або інженерії аналітичних знань:
- два вхідних об”єкта – таблиця кодифікатор та результат розрахунку середніх значень (знання «Основні Статистики»)
- після їх поєднання - результат (знання «Запит» - аналог об’єкта «View» у базі даних)
- останній – кластерно-спектральна карта показників (дивись таблицю 2)

Таблиця 1.

Об’єкт	*Основне призначення*	*Опис*	*Тип результату*
Зовнішня таблиця	Створити вибірку із заданої БД у вигляді таблиці - кодифікатора	Містить початкові дані типу текстового рядка, в яких описані параметри підключення (паролі доступу, місцезнаходження БД, тіло SQL-запиту	Таблиця
«Основні статистики»	Розрахувати основні статистики заданих полів таблиці – середнє, стандартне відхилення тощо	Містить результат розрахунку	Таблиця
«Запит»	Поєднання за кодами	Аналог аналог обєкта «View» у базі даних	Динамічна таблиця
«Кластерно-спектральна карта»	Групування змінних (та випадків) за мірою «схожості»		

Результат розрахунку представлений на рисунку 3.

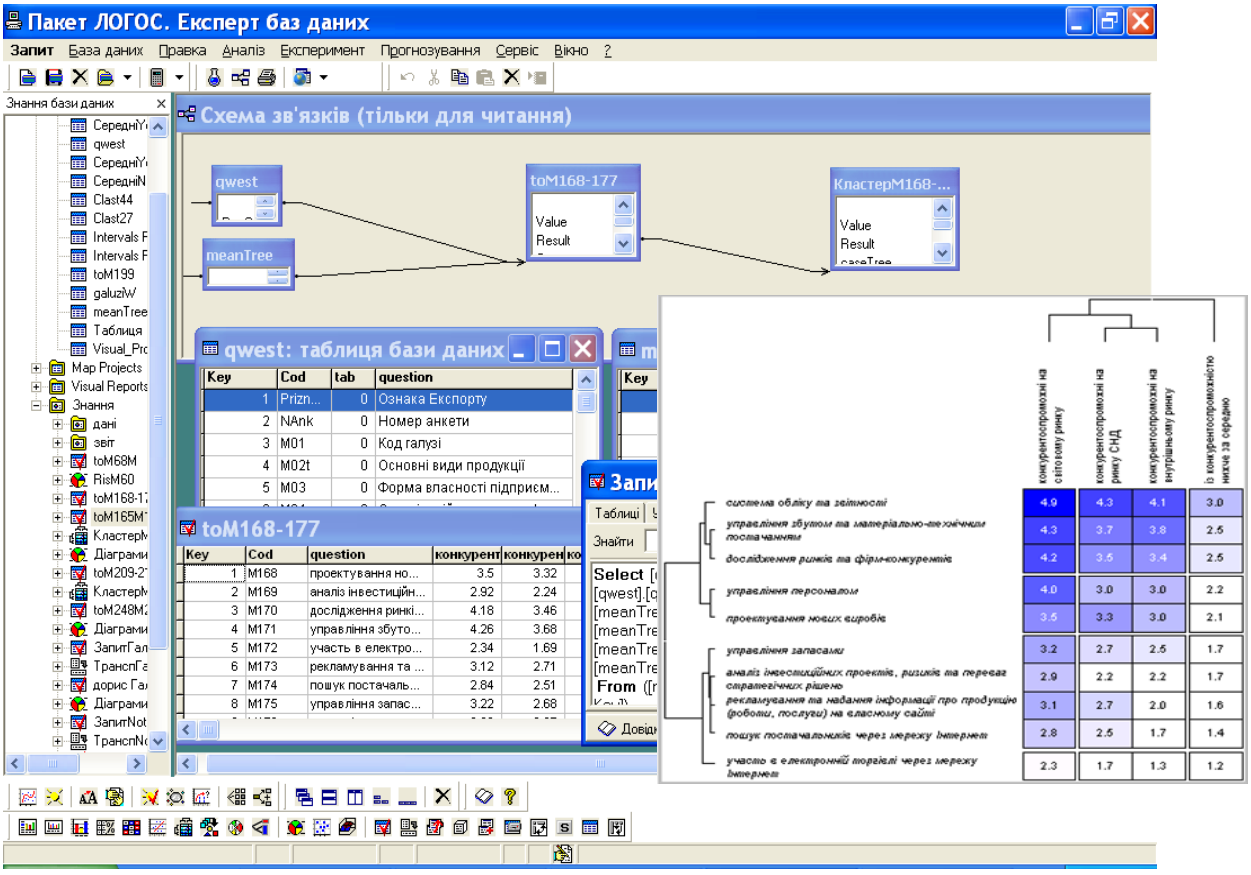

Рис.3. Результат розрахунку ланцюжка методів

Розрахунок виконується автоматично, при чому послідовно виконуються методи, і закінчення виконання попередніх є сигналом для виконання наступних. Таким чином можна відзначити, що така система дозволяє поступове ускладнення методів. Можна одночасно отримувати прогноз поведінки окремих груп (спочатку застосована кластерізація даних, а потім прогноз), можна одночасно отримувати результати розвитку окремих сценаріїв(у вигляді таблиць, після склеювання), обирати найліпший (знаходження оптимуму). Ця система відкрита до поповнення різними об'єктами – методами DataMining, методами прогностики та оптимізації.

Висновки.

Такий підхід дав можливість створювати комплексні результати послідовного застосування аналітичних методів, а це – прогнозування динамічних рядів з одночасним групуванням за тенденціями, це – візуальний факторний аналіз, в якому групуються компоненти за величиною схожості, це – порівняльний розрахунок різних сценаріїв розвитку подій.

Інженерія аналітичних знань перетворюється в побудову інженером гіпотетичної моделі виходячи з системного підходу до аналізу даних, в якій він вибирає за допомогою графічного інтерфейсу методи обробки даних та знань не поглиблюючись в математичні аспекти методу, отримує результат у вигляді аналітичного знання, поступово ускладнюючи ланцюжок обробки знань.

Прикладом реалізації такого підходу є програмне забезпечення інформаційної та експертно-аналітичної підтримки прийняття рішень Логос-3.0, в якому реалізована аналітична база знань у вигляді файлу з розширенням .sdb. База знань має множину об'єктів, що забезпечують роботу з данними та методами пошуку знань.

Список літератури:

1. Барсегян А.А, Анализ данных и процессов: учеб.пособие/ СПб.:БХВ-Петербург, 2009 – 512с.

2. А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. Методы и модели анализа данных: OLAP и Data Mining. – СПб.: БХВ–Петербург, 2004. – 336 с.
3. Сошникова Л.А.,ич В.Н., Уебе Г., Шефер М. Многомерный статистический анализ в экономике. Учеб. Пособие для вузов. – М.:ЮНИТИ-ДАНА, 1999. – 598с.
4. Алгазинов Э.К., Сирота А.А. Системный подход в современных исследованиях. Л. 1978. – 73с.
5. Брандт З. Анализ данных. Статистические и вычислительные методы для научных работников и инженеров. Пер. с англ. – М.: Мир, ООО «Издательство АСТ», 2003. – 686с.
6. Джеймс Одел. Агенти і складні системи.– Ж. «Открытые системы», октябрь 2003. – с.54–58

Відомості про автора:

Лисак Володимир Васильович - старший викладач кафедри Інформаційних технологій Київської державної академії водного транспорту (КДАВТ), аспірант, область інтересів - методи аналізу даних (інтелектуального, чисельного, статистичного тощо)

e-mail: vladimir.lysak@gmail.com