

РЕФЕРАТ

Выпускная квалификационная работа содержит 129 страниц, 29 рисунков, 39 таблиц, 11 источников, 3 приложения.

Ключевые слова: госзакупки, парсер, база данных, OLAP куб, нарушения

Объектом исследования является анализ портала госзакупок.

Цель работы - разработка набора инструментов анализа данных государственных закупок, помогающих аналитикам в нахождении нарушений, а также в формировании статистики.

В процессе исследования использовались методы сбора и анализа информации, методы создания технических документов, методы проектирования и разработки программных средств.

В результате исследования был создан информационно-аналитический программный продукт с использованием разработанных решений на базе технологий многомерного анализа для обнаружения типовых схем нарушений и визуализации информации по госзаказам в удобном для анализа виде.

Программный продукт позволяет актуализировать информацию, получать агрегированные данные в виде отчётов и диаграмм для решения задач получения статистики и выявления нарушений.

Программный продукт востребован предприятиями и организациями, участвующими в госзакупках как в роли заказчика, так и в роли поставщика товаров и услуг, а также контролирующими органами.

Проведено комплексное описание и анализ финансово-экономических аспектов выполненной работы. Рассмотрены вопросы выполнения требований к безопасности и гигиене труда, к промышленной безопасности, охране окружающей среды и ресурсосбережения.

В будущем планируется анализ всех субъектов России, а также использование интеллектуального анализа данных.

ОГЛАВЛЕНИЕ

Введение.....	4
1 Обзор средств для анализа госзакупок	5
1.1 Система государственных закупок	5
1.2 Нарушения в государственных закупках	12
1.3 Обзор аналогов.....	15
1.4 Технологии, используемые при анализе данных	17
1.5 Задание на разработку	18
2 Проектирование средств для анализа госзакупок	20
2.1 Выбор технологий.....	20
2.2 Архитектура системы	22
2.3 Проектирование базы данных	23
2.4 Проектирование парсера	25
2.5 Проектирование OLAP куба	26
3 Разработка средств для анализа госзакупок.....	28
3.1 Реализация ETL-процессов.....	28
3.2 Формирование структуры OLAP куба.....	30
4 Результаты анализа госзакупок	32
4.1 Инструмент эксперта аналитика	32
4.2 Примеры результатов анализа	38
4.3 Нарушения	50
4.3.1 Нарушения со сроками	50
4.3.2 Анализ применения латиницы в русских словах	51

Заключение	54
Список публикаций.....	55
Список используемых источников.....	56
Приложение А	59
Приложение Б.....	68
Приложение В.....	69
Приложение Г	72

ВВЕДЕНИЕ

Использование современных технологий в области организации финансовых отношений между компаниями повлекло к созданию портала государственных закупок Российской Федерации. С 1 января 2011 года портал предоставляет сервисы для поиска информации, размещения заказов на поставки товаров, размещения предложений выполнения работ, оказание услуг для государственных и муниципальных нужд.

В связи с большим объёмом размещаемой информации, а также с несовершенством технических возможностей ресурса, не редко имеют место случаи нарушений среди участников конкурсов.

Целью данной работы является разработка набор инструментов анализа данных государственных закупок, помогающих аналитикам в нахождении нарушений, а также в формировании статистики.

Объектом исследования является анализ портала госзакупок. Инструментами для проведения анализа являются средства технологий On-Line Analytical Processing (OLAP), позволяющие представить информацию в виде, удобном для анализа и находить ранее скрытые знания, например, зависимости между изменениями данных, ранее не рассматриваемых при других видах анализа.

Предметом исследований, представленных в данной работе, является автоматизация процесса сбора и анализа данных портала государственных закупок (сокращенно госзакупки) и предоставление возможности быстрого и гибкого анализа полученных данных. Поскольку объём данных по Российской Федерации слишком большой (около 300Гб), для первичного анализа было решено выбрать одну из областей Российской Федерации. Поскольку работа выполнялась в Томской области, информация о сделках, проводимых в данной области, представляет наибольший интерес.

1 ОБЗОР СРЕДСТВ ДЛЯ АНАЛИЗА ГОСЗАКУПОК

1.1 Система государственных закупок

Портал государственных закупок Российской Федерации [1] — Официальный сайт Российской Федерации в сети «Интернет» для размещения информации о размещении заказов на поставки товаров, выполнение работ, оказание услуг (далее - портал госзакупок). Предназначен для обеспечения свободного и безвозмездного доступа к полной и достоверной информации о контрактной системе в сфере закупок и закупках товаров, работ, услуг, отдельными видами юридических лиц, а также для формирования, обработки и хранения такой информации.

В цели создания портала входит обеспечение прозрачности заключения сделок, оперативное и массовое информирование о предстоящих торгах, общественное обсуждение закупок, повышение конкурентоспособности поставщиков, эффективное расходование государственных средств, и, следовательно, улучшение качества выполнения государственных и муниципальных нужд.

Основными понятиями в сфере госзакупок являются:

- *Государственная закупка* – это процесс приобретения государством в лице уполномоченных органов для обеспечения государственных и муниципальных нужд товаров, работ и услуг.
- *Контракт* - это государственный, муниципальный контракт, гражданско-правовой договор [2].
- В качестве *заказчика* может выступать государственный орган, орган управления государственными внебюджетными фондами, орган местного самоуправления, а также бюджетное учреждение, иной получатель средств федерального бюджета, бюджета Российской Федерации или местного бюджета.

- *Поставщик* — это юридическое или физическое лицо, поставляющее товары/продукты/услуги заказчиком.
- *Контролирующий орган* – орган, следящий за соблюдением порядка в сфере закупок. На федеральном уровне – это федеральная антимонопольная служба России, на региональном и местном уровнях – аналогичные контролирурующие региональные и муниципальные ведомства [3].
- *Закупка* – это совокупность действий, начиная с размещения извещения об осуществлении закупки товара, работы услуги для обеспечения государственных нужд.
- *Объект закупки* – это некая совокупность товаров/продуктов/услуг, одинаковых по содержанию. В одном контракте может быть заказано несколько объектов закупки (далее: продукт).

Процессы государственных закупок в России регулируются Федеральным законом №44-ФЗ «О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд», ранее Федеральным законом №94-ФЗ «О размещении заказов на поставки товаров, выполнение работ, оказание услуг для государственных и муниципальных нужд», а также Федеральным законом Российской Федерации номер №223-ФЗ «О закупках товаров, работ, услуг отдельными видами юридических лиц».

В данной работе анализировались контракты Томской области, публикуемые после 01 января.2014. Контракты заключены по процедуре Федерального закона №44ФЗ, либо №94ФЗ, но имеющих структуру №44ФЗ.

Выделяют следующие принципы закона №44-ФЗ [4]:

- обеспечение гласности и прозрачности в сфере закупок;
- предотвращение коррупции (либо ее минимизация) и других злоупотреблений в сфере закупок;

- повышения качества (эффективности и результативности) обеспечения государственных и муниципальных нужд;
- внедрение системного подхода в сферу планирования и осуществления государственных закупок.

Основные этапы, проводимые на портале госзакупок:

- Размещение заказа
- Приём заявок
- Рассмотрение заявок (проведение конкурса)
- Заключение контракта
- Отслеживание исполнения контракта

Закон номер 44-ФЗ от 05.04.2013 предусматривает следующие аспекты отношений в сфере государственных закупок [1]:

- планирование закупок для государственных и муниципальных нужд;
- размещение извещений о государственных заказах;
- прием конкурсных заявок;
- проведение торгов;
- заключение контракта;
- исполнение контракта.

При планировании закупки, а также подготовке документации и публикации извещения определяется способ размещения заказа (способ определения поставщика). Способ определения поставщика зависит от стоимости закупки, вида объекта закупки, требований к квалификации участников закупки и качественных характеристик и степени сложности в оказании услуг либо поставки товара.

Согласно статье 24 Федерального закона, №44-ФЗ выделяет следующие способы определения поставщика:

- *Открытый конкурс.* Победителем признаётся участник, предложивший лучшие условия выполнения закупки. Кроме

стоимости контракта, оцениваются также качественные характеристики товара, участники закупки и другое.

- *Конкурс с ограниченным участием.* Проводится для закупок, которые в силу их технической и (или) технологической сложности, инновационного или специализированного характера способны выполнить только поставщики, имеющие необходимый уровень квалификации.
- *Двухэтапный конкурс.* Участник должен предквалификационный отбор на первом этапе конкурса и основной отбор на втором этапе. Участник, предложивший лучшие условия исполнения контракта на втором этапе конкурса, объявляется победителем.
- *Закрытый конкурс.* То же, что и открытый конкурс, но информация о закупках сообщается заказчиком путем направления приглашений принять участие в закрытых способах определения поставщиков.
- *Закрытый конкурс с ограниченным участием.* То же, что и открытый конкурс с ограниченным участием, но информация о закупках сообщается заказчиком путём направления приглашений принять участие в закрытых способах определения поставщиков.
- *Закрытый двухэтапный конкурс.* То же, что и открытый конкурс, но информация о закупках сообщается заказчиком путём направления приглашений принять участие в закрытых способах определения поставщиков.
- *Электронный (открытый) аукцион.* Проведение аукционов осуществляется только в электронной форме. Победителем признается лицо, предложившее наиболее низкую цену контракта.
- *Закрытый аукцион.* То же, что и электронный аукцион, но информация о закупках сообщается заказчиком путем направления приглашений принять участие в закрытых способах определения поставщиков.

- *Запрос предложений.* Победителем запроса предложений признается участник закупки, направивший окончательное предложение, которое наилучшим образом соответствует установленным заказчиком требованиям к товару, работе или услуге. В запросе предложений можно указывать только те объекты закупки, которые указаны в ч.2 статьи 83 Закона о контрактной системе.
- *Запрос котировок.* Победителем в проведении запроса котировок признается участник размещения заказа, предложивший наиболее низкую цену контракта. Запрос котировок ограничен ценой контракта либо такими объектами закупки, как продукты, лекарственные средства и другие гуманитарные закупки.
- *Закупка у единственного поставщика.* Не является конкурентной процедурой в отличие от предыдущих процедур. Федеральным законом номер 44-ФЗ предусматривается перечень случаев, когда возможна закупка у единственного поставщика.

Общероссийский официальный сайт (ООС) по государственным закупкам представляет собой единый информационный ресурс, предоставляющий доступ к размещенной информации и располагается по адресу <http://zakupki.gov.ru> (далее – портал закупок).

На портале закупок постоянно пополняющаяся информация об осуществляющихся закупках находится в двух разделах: «Осуществление закупок» и «Информация о контрактах и договорах». Первый раздел пополняется информацией о процессе определения поставщика и стоимости контракта. Во второй раздел выгружается информация о заключённых контрактах (№44ФЗ) и договорах (№223ФЗ). В данной работе анализируется второй раздел.

Ниже показан пример размещения информации о контракте (рисунок 1.1)



ИНФОРМАЦИЯ О КОНТРАКТЕ № 2701731794716000005

ISS - подписка на события контракта ?

Размещено 21.03.2016 12:03 (MSK+3 (UTC+6) Новосибирск)
По местному времени заказчика

ИНФОРМАЦИЯ О КОНТРАКТЕ	ИНФОРМАЦИЯ О ПЛАТЕЖАХ И ОБЪЕКТЕ ЗАКУПКИ	ИНФОРМАЦИЯ ОБ ИСПОЛНЕНИИ (РАСТОРЖЕНИИ) КОНТРАКТА	ДОКУМЕНТЫ	ЖУРНАЛ СОБЫТИЙ
ОБЩАЯ ИНФОРМАЦИЯ				
Реестровый номер контракта	27017317947 16 000005			
Статус контракта	Исполнение			
Номер извещения об осуществлении закупки	0865200000316000055			
Способ определения поставщика (подрядчика, исполнителя)	Электронный аукцион			
Дата подведения результатов определения поставщика (подрядчика, исполнителя)	03.03.2016			
Дата размещения (по местному времени)	21.03.2016 12:03 (МСК+3)			
Реквизиты документа, подтверждающего основание заключения контракта	Протокол подведения итогов электронного аукциона № 0865200000316000055-3 от 04.03.2016; 03.03.2016, №0865200000316000055, протокол подведения итогов электронного аукциона			
ИНФОРМАЦИЯ О ЗАКАЗЧИКЕ				
Полное наименование заказчика	ДЕПАРТАМЕНТ ЛЕСНОГО ХОЗЯЙСТВА ТОМСКОЙ ОБЛАСТИ			
Сокращенное наименование заказчика	Департамент лесного хозяйства Томской области			

Рисунок 1.1. Пример размещения информации о контракте

Информация о контракте включает в себя общую информацию (например, дату заключения контракта, текущий статус, способ определения поставщика), данные о заказчике, поставщике (поставщиках), и объекте(объектах) закупки.

Помимо размещения информации на сайте <http://zakupki.gov.ru/>, портал публикует информацию о сделках на FTP-сервер, содержащий структурированную версию информации сайта. На FTP-сервер регулярно загружается свежая информация о проводимых конкурсах.

Адреса FTP серверов разделены по федеральным законам, применяющимся для размещения тендеров: ФЗ №223; ФЗ №94 и ФЗ №44. Каталоги, содержащие в себе информацию о ФЗ №94 и ФЗ №44, имеют следующую структуру:

- *94fz* каталог, содержащий данные публичных выгрузок в соответствии с 94ФЗ (остальные каталоги - 44ФЗ);
- *fcs_nsi* каталог, содержащий данные нормативно-справочной информации;
- *fcs_regions* каталог, содержащий данные полной региональной выгрузки, опубликованной на ООС информации;
- *fcs_banks* каталог, содержащий данные выгрузки опубликованной на ООС информации по банковским гарантиям;
- *fcs_rules* каталог, содержащий данные выгрузки по правилам (в настоящее время не реализован);
- *fcs_fas* каталог, содержащий данные выгрузки опубликованной на ООС информации о сомнительных поставщиках и проверках;
- *fcs_discussion* каталог, содержащий выгрузку обсуждений

На сервере хранятся заархивированные файлы в формате XML (extensible Markup Language – расширяемый язык разметки), которые содержат информацию об опубликованных документах (извещения, протоколы, контракты и другие). Выгрузка осуществляется в архивируемые файлы формата XML [5]. В одном XML-файле могут находиться документы только одного типа и не больше 3000 записей.

Файлы в выгрузке разделены по каталогам, соответствующим региону. В каждом каталоге региона есть также каталоги: SketchPlans, PlanGraphs, Notifications, PurchaseDocs, Protocols, Contracts, AddInfo, CustomerReports, RegulationRules, RequestQuotation, AuditResult. В каждом из каталогов дополнительно есть каталоги currMonth и prevMonth.

Структура папки *44fz* показана на рисунке 1.2.

–	<Наименование региона>;
–	notifications;
–	currMonth;
–	prevMonth;
–	protocols;

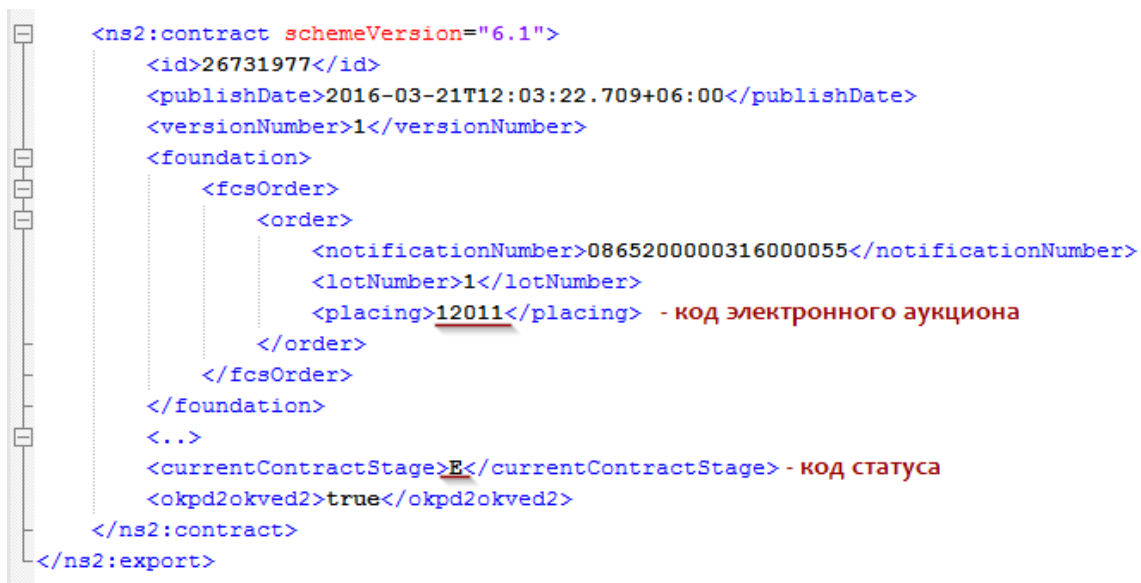
```

        - currMonth;
        - prevMonth;
    - contracts;
        - currMonth;
        - prevMonth;
    - <...>
- < Наименование региона >;

```

Рисунок 1.2. Структура папки 44fz на FTP-сервере сайта о государственных закупках

Данные по заключённым контрактам находятся в каталоге *fcs_regions/Tomskaja_obl/contracts/*. Пример XML файла, находящегося в каталоге и являющегося выгрузкой контракта, представленного на рис1.1, показан ниже (рисунок 1.3)



```

<ns2:contract schemeVersion="6.1">
  <id>26731977</id>
  <publishDate>2016-03-21T12:03:22.709+06:00</publishDate>
  <versionNumber>1</versionNumber>
  <foundation>
    <fcsOrder>
      <order>
        <notificationNumber>0865200000316000055</notificationNumber>
        <lotNumber>1</lotNumber>
        <placing>12011</placing> - код электронного аукциона
      </order>
    </fcsOrder>
  </foundation>
  <...>
  <currentContractStage>E</currentContractStage> - код статуса
  <okpd2okved2>true</okpd2okved2>
</ns2:contract>
</ns2:export>

```

Рисунок 1.3. Аналог общей информации в XML файле

1.2 Нарушения в государственных закупках

Поскольку технические возможности ресурса не совершенны, не редко имеют место случаи нарушений среди участников конкурсов. Соответственно есть необходимость в методологиях и технологиях, помогающих в выявлении случаев нарушений.

Встречаются махинации, как и проводимые с помощью торговой площадки, так и вне ее, например, подделка документов и выдача

конфиденциальной информации конкурентам. Операции нарушений, которые фиксирует система закупок, можно отследить при помощи анализа информации с сайта, тогда как шантаж и выдача конфиденциальной информации выявляются совершенно другими методами. Далее описаны самые распространенные шаблоны нарушений [6].

Размещение закупок с избыточными потребительскими свойствами. Обычно выражается в закупке очень дорогих товаров у проверенных поставщиков. При заключении такой сделки присутствуют необоснованные траты бюджета, расточительство и возможное ограничение конкуренции.

Заключение контракта не с победителем. Заказчик заключает контракт не с тем участником, который предложил минимальную стоимость контракта, а с тем, кто предложил больше денег. Среди последствий такой схемы можно выделить завышение общей цены контракта, ограничение конкуренции, нерациональные бюджетные траты.

Разбивка заказа на несколько отдельных процедур. Результатом сговора сотрудников организаций может быть решение разбить большую закупку на несколько небольших заказов, чтобы избежать проверки, которой подвергаются крупные закупки. Сутью схемы является сокращение минимального срока уведомления для ограничения конкуренции в пользу компаний, которые были осведомлены о предстоящих торгах

Сильное сокращение срока исполнения заказа. Заказчик указывает невыполнимые сроки исполнения контракта в техническом задании или проектно-сметной документации. Таким образом, установленный документацией объем работ за указанный срок сможет выполнить только тот участник, который уже сделал часть работы, предварительно начав исполнение еще до выставления заказа. Обнаружение заключается в поиске контрактов с датой исполнения, близкой к дате окончания процедур закупки, и сопоставлении сроков с объемом работ.

Затруднение поиска заказа в базах. Иногда одна или несколько букв в названии тендера пишется латиницей или просто допускаются опечатки в ключевых словах. Поставщики не могут обнаружить данный заказ, потому что при поиске набирают корректный вариант. В эту же категорию можно отнести заказы, в название которых ставят лишние пробелы, тире или нижние подчеркивания. Также для каждой услуги или товара есть свой код ОКПД, по которому и ориентируются компании-поставщики. Если услуги по строительству поместить в графу «Сантехника», строительные компании не смогут их найти. Смысл этих ухищрений один – найти объявление о заказе смогут только заранее оповещённые граждане.

Конкретизация технического задания. Невыполнимые требования могут предъявляться не только к реализации государственного контракта, но и к претендентам. Особенно часто такой способ используют при закупках лекарств.

Большая часть схем основывается на сговоре участников процедур закупок. Это может быть сговор между участниками закупок или между участниками закупок и организаторами закупок.

Для обнаружения сомнительных сделок по некоторым перечисленным схемам потребуется анализ документов закупки, например, технического задания, проектно-сметной документации и других. Такая информация достаточно разнородна и хранится в файлах различных форматов. Синтаксический анализ этих документов увеличит время извлечения информации и может не привести в итоге к ожидаемому результату, так как распознавание и анализ текста является отдельным направлением в области искусственного интеллекта.

Для выявления сомнительных сделок с помощью методов интеллектуального анализа данных наиболее подходят ситуации с закупками по завышенным ценам и предметов роскоши, заключением контракта не с победителем аукциона, демпингом цен на торгах, нереальными сроками исполнения контракта, двойным проведением закупки, проведением конкурса

вместо аукциона и разбиением заказа на несколько маленьких лотов. Все необходимые данные по этим схемам находятся в извещениях о закупках, подписанных контрактах и протоколах комиссии.

Нарушители зачастую действуют по каким-то известным схемам. Зная какую-то схему, можно применить эвристический анализ для распознавания подобных ситуаций и их предотвращения. Но, анализируя большое количество сделок, гораздо эффективней будет применять автоматизированные методы анализа данных.

1.3 Обзор аналогов

Портал государственных закупок отслеживает взаимодействие огромного количества государственных и частных компаний. Портал также является ресурсом, на котором данные с различных торговых площадок появляются в первую очередь (практически моментально). Эти факторы побуждают множество компаний следить за обновлениями сведений о закупках на портале. Однако ресурс имеет негибкую форму поиска данных, на портале отсутствует сравнение конкурентов, что является некомфортным для пользователя. Поэтому, исходя из различных целей, сторонние разработчики выпускают продукты, помогающие пользователям эффективно работать с порталом.

В связи со спросом на рынке уже существуют различные системы для работы с сайтом закупок, которые ориентированы в первую очередь на поставщиков, удобство и точность поиска, фильтрацию извещений о размещении заказов. Такие системы практически не предлагают аналитики или инструментов для проведения анализа данных государственных закупок.

1. **Закупки 360** — это платная онлайн-система мониторинга и анализа государственных и коммерческих закупок, публикуемых на 69 электронных торговых площадках, позволяющая находить и отслеживать подходящие закупки, активности заказчиков, конкурентов,

победителей гос. и коммерческого заказа с целью увеличения собственных продаж [7]. Сервис также предлагает статистические и аналитические услуги: просмотр статистики побед поставщиков у заказчика, прогнозирование присутствия участников и вероятных цен, риски срыва сроков подготовки заявки.

2. **ГосЗакупки** — это платный комплекс программ, которые являются альтернативным вариантом поиска аукционов, котировок, тендеров и контрактов [8]. Позволяет получить данные в виде таблиц в одном окне. Обладает высокой скоростью поиска информации. К недостаткам сервиса относится отсутствие визуального представления данных и отсутствие аналитики.

Помимо систем, рассчитанных на участников конкурсов, существуют решения, ориентированные на обеспечение прозрачности заключения сделок, повышение общественной осведомлённости, контроль закупок и выявление нарушений в сфере государственных закупок. Одним из таких решений является проект Госзатраты.

3. **Госзатраты** - это автоматическая система мониторинга портала госзакупок. Сайт постоянно обновляет информацию о контрактах, закупках, заказчиках и поставщиках, однако поисковая система неудобна для оперативного поиска нужных тендеров, так как проект больше ориентирован на выявление нарушений и повышение общественной осведомлённости в области расходования государственных средств [9]. Так, Госзатраты автоматически находят такие виды нарушений, как неполную или искажённую информацию об объекте закупки, закупку предметов роскоши, контракты с запрещёнными или ограниченными объектами закупки, заниженные сроки исполнения контракта и наличие поставщика в реестре недобросовестных поставщиков. Но, несмотря на автоматический анализ, найденные данные нуждаются в перепроверке. На сайте также

присутствует раздел визуализации, однако данные в этом разделе являются неактуальными, так как в последний раз обновлялись в начале 2015 года. Помимо средств автоматического выявления нарушений, на сайте присутствует раздел Новости, где регулярно публикуются подозрительные закупки на крупные суммы и другие новости, касающиеся сферы Госзакупок.

Ни одна из рассмотренных систем не предоставляет широкие возможности для анализа данных портала госзакупок. Существующие разработки имеют ограниченные функциональные возможности по анализу: в системах отсутствует либо строго прописан анализ данных. Так, на Закупках 360 анализ привязан к заказчику или объекту закупки, а на Госзатратах происходит выявление только семи видов нарушений.

1.4 Технологии, используемые при анализе данных

При решении задач анализа данных обычно приходится собирать данные из нескольких источников различной структуры. Основная проблема при этом состоит в несогласованности и противоречивости этих источников, отсутствии единого логического взгляда на обрабатываемые данные. Решением этой проблемы является использование хранилища данных.

Хранилище данных (DataWarehouse, DWH) представляет собой специально организованную единую базу данных предприятия, в которой обеспечивается сбор, хранение и быстрый доступ к предметно-ориентированной, интегрированной и поддерживающей хронологию корпоративной информации [10].

Сбор данных в корпоративное хранилище производится с помощью ETL-процессов (от англ. extract, transform, load – дословно «извлечение, преобразование, загрузка»), обеспечивающих не только перенос данных в хранилище из различных систем-источников, но и их трансформацию в соответствии с заданными правилами, устранение противоречий.

Архитектура хранилища данных состоит из трёх компонентов:

- источник данных: может содержать структурированные данные в виде таблиц, файлов, документов различных форматов;
- промежуточная область: содержит вспомогательные таблицы, создаваемые для организации процесса выгрузки;
- получатель данных: хранилище данных или база данных, в которую должны быть помещены извлечённые данные.

Часто с хранилищами данных используется и технология OLAP (On-Line Analytical Processing – аналитическая обработка в реальном времени) – технология, которая позволяет пользователям многомерных баз данных генерировать в режиме реального времени наглядные и сравнительные итоговые отчеты по данным.

Для решения задач анализа данных решено построить OLAP-куб, т.е. многомерную базу данных.

К задачам анализа собранной информации относятся: сравнительный анализ стоимости услуг и товаров (переплачивает ли заказчик), сравнение цен на продукты и услуги в зависимости от времён года, от региона, доля местных и внешних к региону поставщиков, визуализация общей статистики в виде графиков, диаграмм.

1.5 Задание на разработку

В рамках магистерской диссертации была поставлена цель: разработать набор инструментов анализа данных государственных закупок, помогающих аналитикам в нахождении нарушений, а также в формировании статистики.

Для достижения цели были поставлены следующие задачи:

- собрать и проанализировать научно-техническую информацию по тематике исследования;
- провести анализ существующих решений;
- проанализировать типовые схемы нарушений;

- реализовать структуру базы данных;
- разработать архитектуру информационно-аналитической системы;
- спроектировать и реализовать ETL процессы;
- построить OLAP куб для проведения многомерного анализа;
- провести анализ полученных данных
- выполнить поиск нарушений

Разработка ведётся с применением технологий извлечения, преобразования и загрузки данных, реализации хранилища данных, многомерного анализа данных. В качестве подсистем для реализации установлены следующие: подсистема для управления ETL процессами, хранилище данных, подсистема для анализа данных, инструмент для поиска отдельных видов нарушений.

2 ПРОЕКТИРОВАНИЕ СРЕДСТВ ДЛЯ АНАЛИЗА ГОСЗАКУПОК

2.1 Выбор технологий

Выбор технологий системы зависит от нескольких факторов: время, отведённое на проект, требования к используемым продуктам разработки, потребности пользователей, требования к сопровождению и стоимости.

На предприятии ООО «СибХайТекЦентр» активно используются продукты разработки ПО компании Microsoft. Соответственно, требованием заказчика являлось разработать систему, которая будет поддерживаться с использованием продуктов этой компании.

Инструменты Microsoft позволяют минимизировать временные затраты на создание хранилищ данных и делают процесс анализа доступным для пользователя. Цель продуктов Microsoft состоит в том, чтобы упростить разработку, внедрение и администрирование решений на основе хранилищ данных.

Microsoft предлагает комплексное решение по созданию хранилищ данных (см. рисунок 2.1). Оно предполагает использование данных внешних источников, которые преобразуются и заносятся в базу данных (далее БД). Далее с помощью службы Analysis Services данные преобразовываются в многомерный OLAP-куб, который можно анализировать различными средствами: Analysis Services, MS Office и другими.

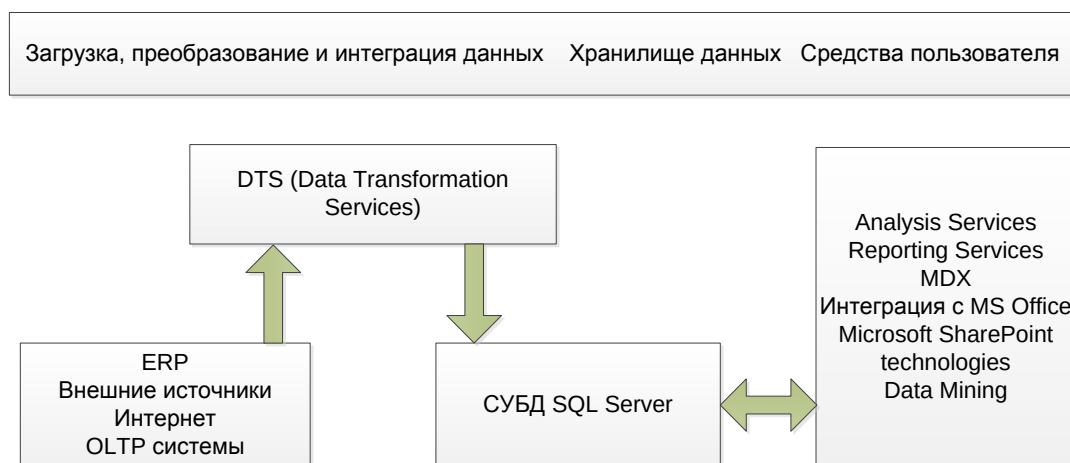


Рисунок 2.1. Комплексное решение Microsoft по созданию хранилищ

Чтобы не ограничивать возможности разрабатываемой системы по выбору СУБД, было решено заносить данные в БД с использованием языка программирования Python. Следовательно, в схеме, представленной на рисунке 2.2, продукт DTS Integration Services, заменён на скрипты, написанные на Python.

Для разработки синтаксического анализатора данных (далее: парсер) было принято решение использовать библиотеки ElementTree – библиотека для разработки парсера, и pyodbc – библиотека, позволяющая взаимодействовать с БД MS SQL, входящие в состав Python 3.3, для получения первичных данных. С помощью данных библиотек реализуется парсер, способный извлекать информацию из ресурса и заносить её в БД.

Для хранения полученных данных используется система управления базами данных Microsoft SQL Server 2012. Она обладает высокой скоростью обработки данных, дополняется технологиями анализа больших данных, а также входит в комплексное решение Microsoft по созданию хранилищ данных.

Для создания многомерных моделей и интеллектуального анализа данных в SQL Server и Visual Studio 2015 (при условии использования дополнения Business Intelligence) использовались службы Analysis Services. Эта подсистема предоставляет аналитические данные, которые применяются в бизнес-отчетах и клиентских приложениях, таких как Excel.

Для анализа готового куба было принято решение использовать Microsoft Excel 2013 Professional +. Поскольку Excel является продуктом компании Microsoft, он позволяет тесно интегрироваться со службами Analysis Services. Кроме того, MS Excel 2013 Professional + имеет мощный механизм визуализации данных, позволяющий строить различные диаграммы, а также отображать данные на карте.

Итоговое решение, основанное на технологиях Microsoft, представлено на рисунке 2.2

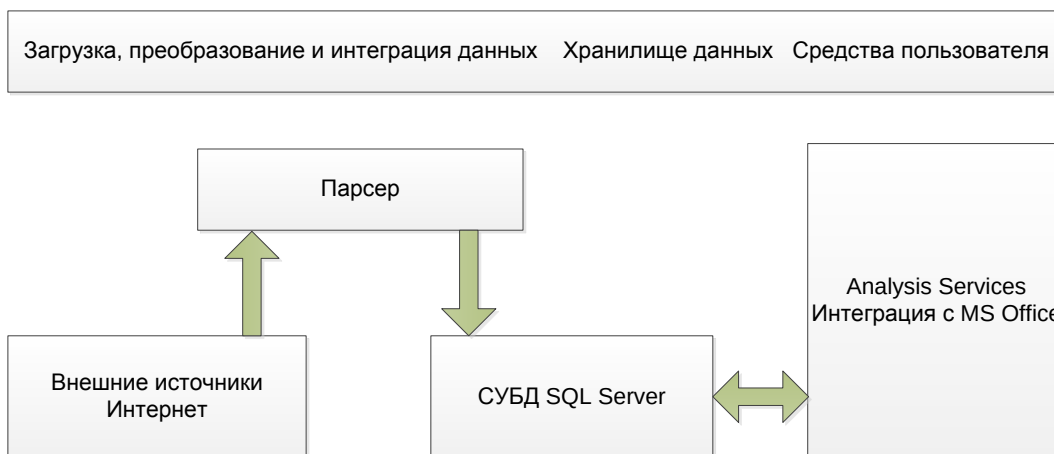


Рисунок 2.2. Архитектура набора по созданию хранилищ

Преимущества использования данных технологий:

- Все бизнес-направления поддерживаются за счет готовых сервисов.
- Такие решения хороши для создания хранилища данных, которое предполагается использовать длительное время.
- Такие решения подходят для сложной аналитической обработки данных, требующей разработки специальных приложений для анализа.
- Данные технологии имеют широкое распространение, следовательно, существует достаточное количество практических материалов, книг, позволяющих быстро и качественно освоить данные инструменты.

2.2 Архитектура системы

Архитектура системы хранения данных корпоративного хранилища состоит из следующих областей:

- *Источники данных* – внешние источники, данные которых необходимо анализировать.
- *Область временного хранения данных* - предназначена для временного хранения данных, извлеченных из систем-источников; является промежуточным слоем между источниками данных и хранилищем данных

- *Постоянное хранилище данных*, которое включает витрины данных – срез хранилища данных, представляющий собой массив тематической, узконаправленной информации, ориентированный, например, на конкретную область или задачу. Витрины реализуются по схеме «звезда», «снежинка», «многие-ко-многим» и содержат выборку данных из хранилища данных. Они также могут быть реализованы в виде многомерного OLAP-куба [11].
- *Модуль аналитики* предназначен для анализа данных, получаемых составлением срезов витрин.

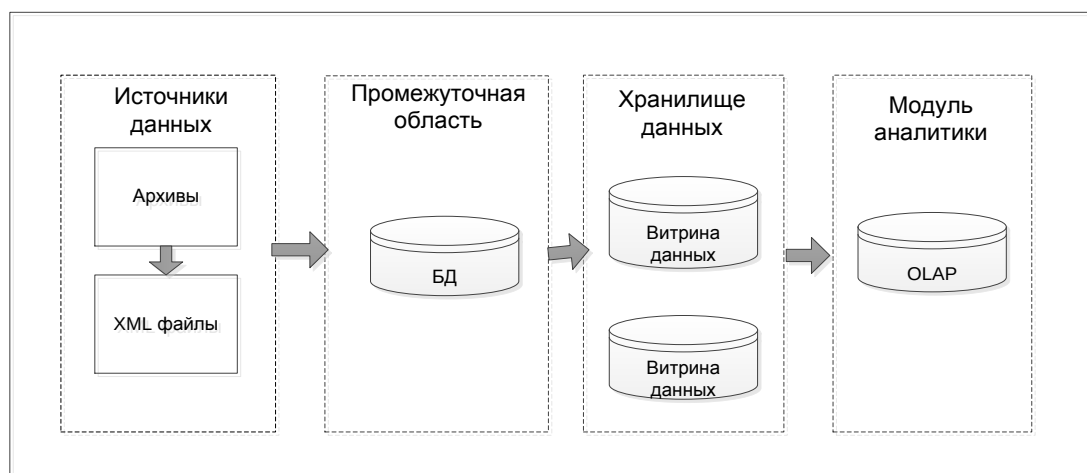


Рисунок 2.3. Архитектура системы

2.3 Проектирование базы данных

Данные с источников загружаются в реляционную базу данных. Для формирования БД было необходимо проанализировать источник данных, которым является FTP сервер портала госзакупок.

Пример упрощённого XML файла представлен в приложении Г. Структура документа имеет сложные типы объектов и большое количество вложенных элементов. Это представляет некоторую сложность при извлечении и синтаксическом анализе данных из XML документов. Описание схемы XML файлов можно найти на портале государственных закупок в документе Альбом

ТФФ ПУ РК РБГ_6.2 [5], также там располагаются схемы XML данных, используемые при выгрузке данных на FTP сервер. При проектировании базы данных использовались части схемы *zfps_contractType*, *zfps_contract2015Type* и вложенные структуры. Пример XML файла представлен в приложении Г.

В элементе *export* содержится информация об используемой схеме, включая версию схемы.

По результатам анализа источника данных, в частности структуры XML файлов, была спроектирована база данных. Рассмотрим, часть БД, включающую таблицы, которые содержат основные сведения о заключении контрактов (рисунок 2.4):

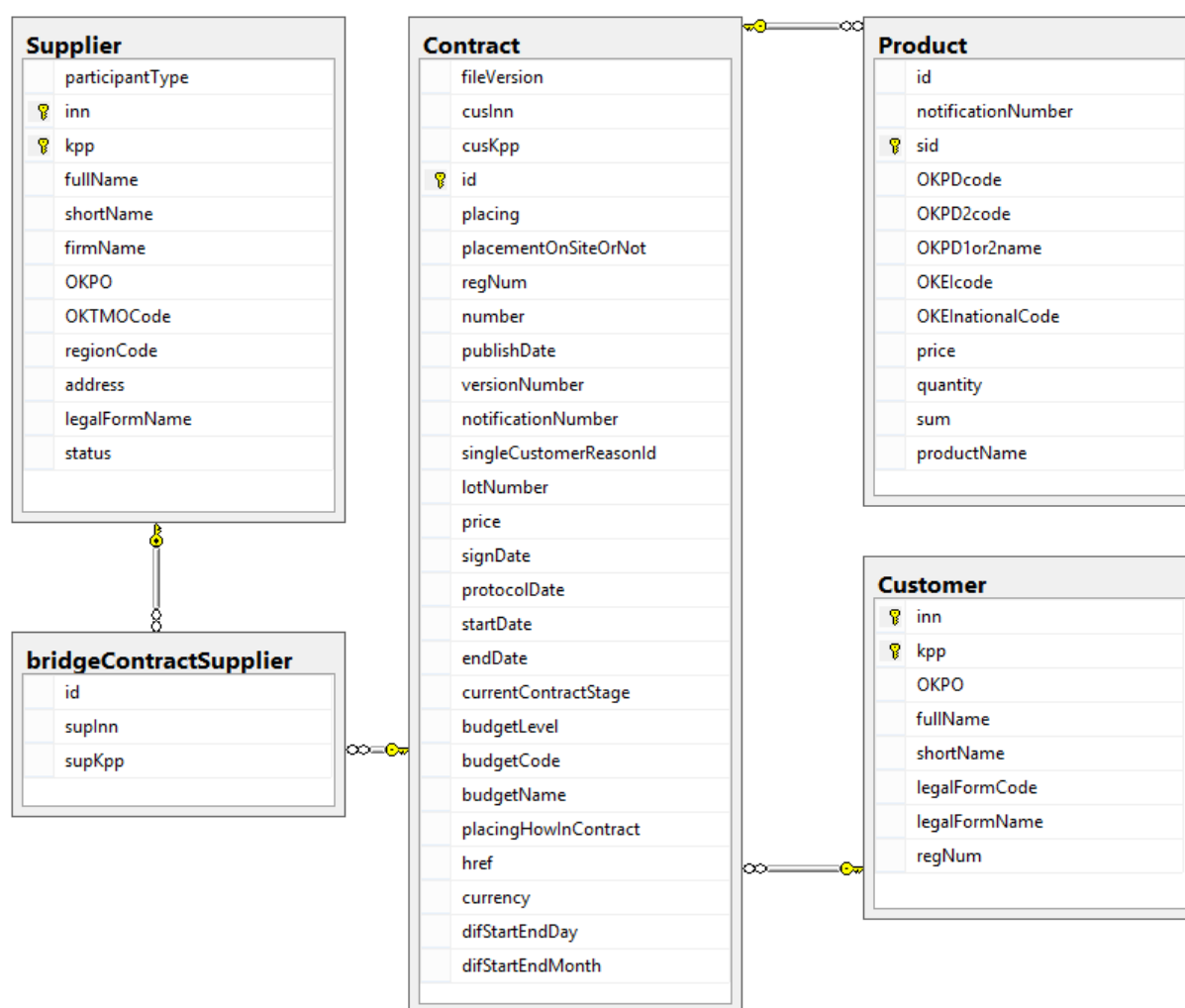


Рисунок 2.4. Таблицы, содержащие основные сведения о заключении контрактов

Среди основных элементов, заносящихся в БД, можно выделить следующие: идентификационные данные контракта, заказчика, поставщика и продуктов, версия контракта, способ размещения (см. приложение В), данные о *заказчике* - имя, *поставщиках*: имя, контактные данные, тип поставщика: юридическое лицо, индивидуальный предприниматель или физическое лицо; правовая форма, адрес, название предприятия.

Учитываются также данные об *объекте закупки*, такие как код продукта – ОКПД1/2, название, единицы измерения заказа: ОКЕИ код, цена за ед. товара, количество товара и общая сумма. Код ОКПД помогает унифицировать виды закупок, он широко используется при анализе данных. В данных о *контракте* учитываются, помимо вышеперечисленного, цена, состояние контракта, используемый бюджет (например, федеральный, муниципальный, бюджет организации), период выполнения контракта, дата (подписания) протокола, дата заключения контракта и ссылка на страницу закупки на официальном сайте.

Полная схема БД приведена в приложении Б.

2.4 Проектирование парсера

При разработке аналитической системы с организацией хранилища данных необходимо, чтобы хранилище было заполнено актуальной для анализа информацией и постоянно пополнялось. Отсюда следует необходимость перемещения данных из систем оперативной регистрации данных в хранилище. Основными этапами этого перемещения являются: извлечение, преобразование и загрузка или ETL-процесс. ETL-процесс включает в себя:

- извлечение данных из внешних источников;
- их трансформация и очистка для соответствия потребностям бизнес-модели;
- и загрузка их в хранилище данных [12]

В данной работе ETL-процесс выполнялся парсером, написанном на языке Python. Обобщенная схема парсера для извлечения, преобразования и загрузки данных по государственным закупкам представлена на рисунке 2.5.

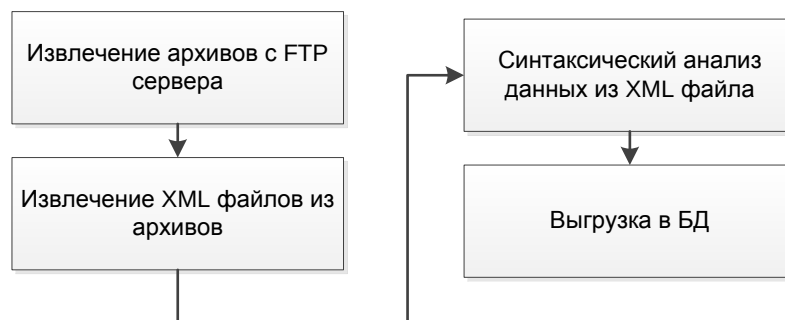


Рисунок 2.5. Схема работы парсера

При синтаксическом анализе данных из XML файлов считывается необходимая информация, при необходимости редактируется и заносится в БД. Парсер предусматривает случаи отсутствия данных или некорректного заполнения значений.

Загрузка данных в БД происходит последовательно. Порядок заполнения зависит от связей в БД: сначала необходимо заполнить таблицы-справочники, затем таблицы *Supplier*, *Customer*, *Contract*, *bridgeContractSupplier*, *Product*.

2.5 Проектирование OLAP куба

В БД содержатся данные, организованные согласно реляционной модели хранения данных. Схема базы является нормализованной и содержит множество взаимосвязанных таблиц. Поэтому выполнение сложных запросов приводит к объединению многих таблиц, что существенно увеличивает время отклика системы. Проектирование хранилища данных подразумевает создание денормализованной структуры данных, ориентированной в первую очередь на высокую производительность при выполнении аналитических запросов.

Кубы - это средства быстрого доступа к данным, содержащимися в большом хранилище данных. Они организуют и сводят данные в многомерную

структуру. Кубы являются основным компонентом технологии OLAP и предоставляют простой в использовании механизм для запроса данных с малым и предсказуемым временем ответа. Куб состоит из данных измерения и мер.

В кубе используются схемы связей: «звезда» и «многие-ко-многим». Модель данных типа «звезда» состоит из двух типов таблиц: в центре «звезды» находится таблица мер, и нескольких таблиц измерений – лучей звезды.

Таблицы мер дают числовую характеристику какому-то аспекту предметной области и содержат одну или несколько колонок вычисляемого типа. В магистерской диссертации мерами куба являются число заказчиков, поставщиков, контрактов, продуктов, цена контракта, продукта и др. (Приложение В).

Измерения – это определяющие атрибуты фактов, и обычно отвечают на вопросы: когда произошел факт, над чем или с чем именно, кто был объектом или субъектом и т.д. В дипломной работе измерениями являются заказчики, поставщики, продукты, контракты, время заключения контракта.

Обычно данные в таблицах-измерениях денормализованы: ценой несколько неэффективного использования дискового пространства удается уменьшить число участвующих в операции соединения таблиц, что обычно приводит к сильному уменьшению времени выполнения запроса [11].

В БД присутствует связь многие ко многим, которую нельзя добавить в модель данных «звезда». В таком случае требуется устанавливать дополнительную меру при разработке куба, которая в базе данных является промежуточной таблицей. Подобную модель данных следует использовать, только если невозможно применить стандартные схемы связей «звезда» и «снежинка», так как они отрицательно сказываются на скорости обработки запросов.

3 РАЗРАБОТКА СРЕДСТВ ДЛЯ АНАЛИЗА ГОСЗАКУПОК

3.1 Реализация ETL-процессов

Парсер представляет из себя совокупность скриптов, скачивающих, анализирующих данные и заносящих их в БД. В парсере прописывается, какую информацию надо занести, предупреждается возможное наличие в файлах некорректных значений либо их отсутствие.

В Приложении Г приведён код скрипта, использующейся для загрузки файлов с удаленного FTP сервера на локальную машину пользователя, а также для распаковки архивов. Пример кода скрипта представлен.

Скрипт проверяет, обрабатывался ли архив ранее. Для этого в базе данных существует вспомогательная таблица, которая содержит в себе названия архивов, которые уже были анализированы. Если архива в списке нет, он скачивается для разархивирования.

После загрузки файлов выполняется процесс синтаксического анализа. В данной части программы происходит подключение библиотеки *lxml*, созданной для более удобного и быстрого парсинга XML, и *pyodbc* для работы с БД. Для работы со значениями времени используется библиотека *datetime*. В процессе работы программа получает название файла, который необходимо проанализировать, задаёт стандарты, использующиеся при создании файла. В файлах используется пространство имён, поэтому при получении данных необходимо это учитывать.

Часть кода скрипта, парсящего файлы для заполнения таблицы *Контракты*, представлена в приложении Г.

Для использования сравнений строк при занесении в БД вместо NULL при отсутствии значений записывается «-». Т.к. в формат даты нельзя записать подобное значение, то выбрано единое значение '1800-01-01', которое потом при помощи вспомогательного скрипта заменяется на NULL.

Структура скриптов преобразования и занесения в БД данных использует одинаковую логику. В начале объявляются переменные. Это сделано для предупреждения отсутствия данных. Под отсутствием данных понимается отсутствие тега и значения в файле XML.

Далее открывается файл XML и происходит построчное записывание данных в переменные. Т.к. структура файла является вложенной, а положение значений во вложенности в зависимости от версии файла разное, то строки, парсящие одно и то же значение, могут встречаться на разных уровнях сложности по 2 раза. Основную сложность разработки парсера составляла разработка алгоритма, при котором данные при наличии в одном контракте нескольких объектов закупки либо поставщиков, записывались корректно. Поэтому поиск значений происходит строго по тегам в определённом цикле. Итеративный цикл не используется для избегания сдвига данных.

После того, как заполнились переменные, происходит запись строки в таблицу БД. При записи происходит проверка наличия такой же строки в таблице. Эта проверка делается для того, чтобы в случае повторного запуска скрипта данные не записывались повторно и не возникало ошибки дублирования ключа.

В результате применения синтаксического анализатора была сформирована база данных контрактов Томской области. Размер БД составляет 4,5 Гб. В ней содержится 17 таблиц, включая 4 таблицы с основными данными, 1 вспомогательную таблицу для связи поставщиков и контрактов, 11 таблиц справочников и 1 таблицу с результатами выявления нарушений. В базе данных содержатся сведения о 1 227 заказчиках, 8 257 поставщиках, 56 608 заключённых контрактах и 247 406 объектах закупки.

3.2 Формирование структуры OLAP куба

Разработка куба выполнялась с помощью средств, предоставляемых службами Analysis Services, и дополнением Business Intelligence в Visual Studio 2015.

После соединения с источником данных было необходимо сделать вычисляемые столбцы. При наличии составного ключа необходимо создание вычисляемого столбца, куда будут записано общее значение составного ключа. В дальнейшем составной ключ используется для создания связей в кубе. Вычисляемые столбцы также удобно делать для общей записи фамилии, имени и отчества. Вычисляемые столбцы не сохраняются в БД. Меры добавляются в процессе или после создания куба.

Были сформированы следующие измерения:

- *Поставщики* – содержат агрегированные данные об исполнителях
- *Продукты_контракты_заказчики* - измерение, содержащие в себе агрегированную информацию о продуктах, контрактах и заказчиках. Данные совмещены с целью быстрой обработки информации.
- *Время* – измерение времени. Привязывается к атрибуту измерения «Дата заключения контракта».

Измерение *Поставщики* нельзя занести в общее измерение, так как связь, соединяющая таблицу поставщиков с таблицей контрактов – многие ко многим. Для этого в меры куба добавляется связывающая таблица – *bridgeContractSupplier*, которая является мостом для взаимодействия мер и измерений, связанных с исполнителями. Измерения *Время* привязано к дате заключения контракта.

Измерения и группы мер связываются между собой с помощью уникальных идентификаторов. В кубе присутствуют связь многие ко многим, обычная – для привязки измерения времени, фактическая – для привязки мер к измерениям, чьи таблицы содержат эти меры. В окне куба «Использование

измерений» задаются связи между мерами и измерениями, т.е. осуществляется формирование модели куба.

Измерения		Группы мер			
		Contract	Product	Supplier	Bridge Contract Sup...
Продукты_контракты_заказч...	Id	Sid	Bridge Contract Su...	Id	Inn
Поставщики (Supplier)	Bridge Contract Su...	Bridge Contract Su...	Inn	Inn	Contract
Время	Дата	Contract	Bridge Contract Su...	Contract	Contract

Рисунок 3.1. Создание связей куба

Во вкладке браузер находится представление куба: в левой части список мер и измерений, в верхней части фильтр, в центре выводятся результаты. Однако для вложенных измерений удобней работать в MS Excel. Ссылка на него находится на панели инструментов браузера.

Измерение Поставщики содержит 10 атрибутов, измерение *Продукты_контракты_заказчики* - 30 атрибутов. Диапазон времени содержит значения с 1.01.1999 до 31.12.2016. Количество мер куба - 10. Полное описание мер и измерений представлены в приложении В.

4 РЕЗУЛЬТАТЫ АНАЛИЗА ГОСЗАКУПОК

4.1 Инструмент эксперта аналитика

Благодаря использованию продуктов Microsoft при разработке, существует возможность пользования лёгким для понимания, но в то же время мощным инструментом анализа данных – MS Excel 2013.

Разработанный программный продукт представляет широкие возможности для анализа данных портала госзакупок. Для исследований достаточно знать предметную область. Представляется возможность сравнения и агрегации атрибутов между собой. Инструмент позволяет анализировать конкретного участника торгов либо продукт; данные, относящиеся к группе участников/продуктов (например, вид бюджета, тип ОКПД), составлять аккумулируемые данные в виде таблиц, наглядные показатели - диаграммы. Использование надстройки MS Excel - Power View, даёт возможность просмотра данных по регионам на карте.

В приложении В показан полный перечень сущностей и атрибутов, по которым возможно проводить сравнения, а также мер.

На рисунке 4.1 показано окно анализа:

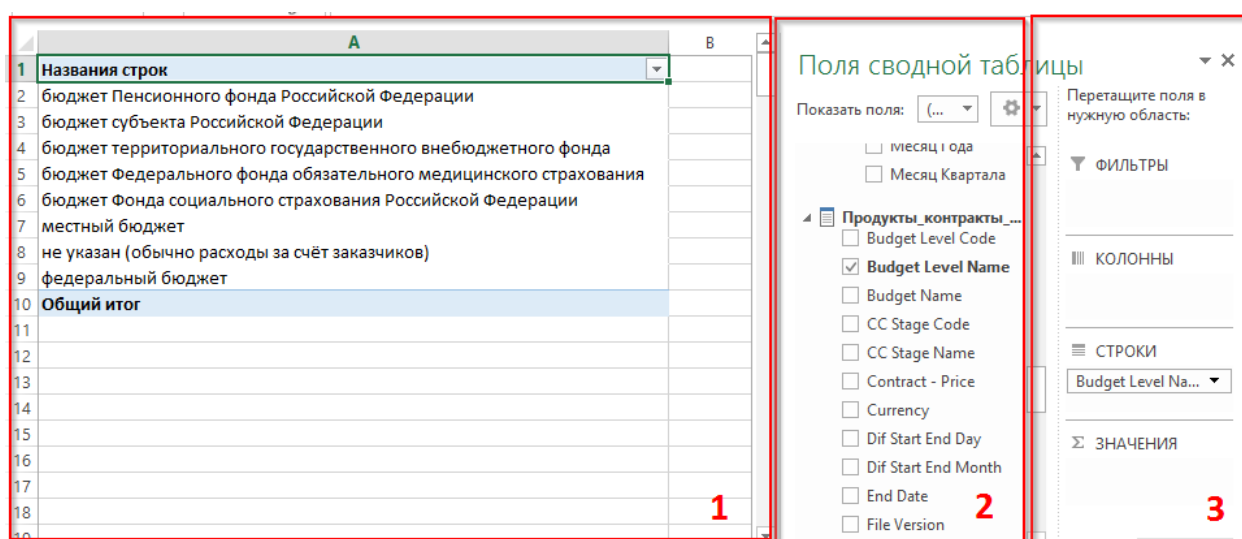


Рисунок 4.1. Окно анализа в MS Excel

На данном рисунке представлены следующие области:

- Область выдачи результатов (1)
- Область мер и измерений (2)
- Область указания/назначения структуры среза (3)

Составлять срезы куба возможно, используя следующие параметры:

- *Фильтры* – будут показаны только те данные, которые подходят параметрам
- *Строки* – данные будут отображаться в строках
- *Колонны* – данные будут отображаться в колоннах
- *Значения* – область, где указываются, в чём будет измеряться срез.

Разработанный набор программных средств позволяет за короткий срок получить агрегированную информацию об интересующих данных. В качестве примера на рисунке 4.2 показана структура среза, а на рисунке 4.3 результаты, оформленные в виде таблицы и отражающие продажи крупнейших индивидуальных предпринимателей и физических лиц Томской области за 2015 год.

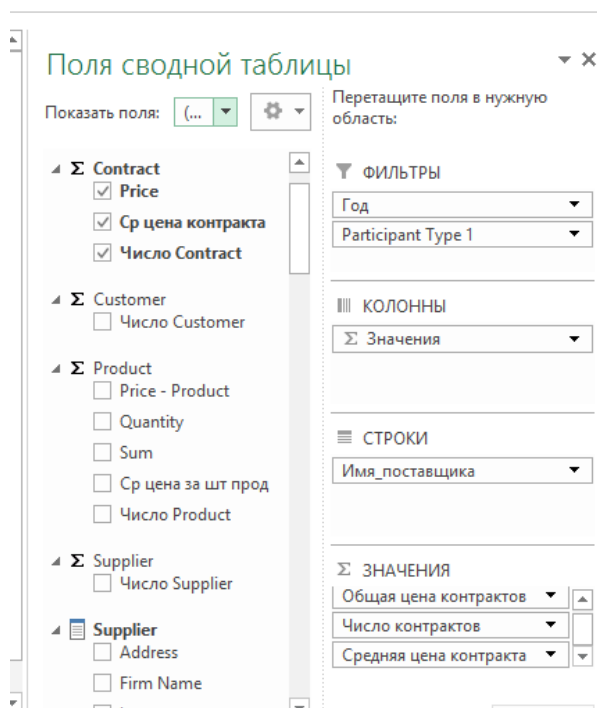


Рисунок 4.2. Пример структуры среза

	A	B	C	D
1	Год	Calendar 2015		
2	Participant Type 1	(несколько элементов)		
3				
4	Названия строк	Общая цена контрактов	Число контрактов	Средняя цена контракта
5	Шакуло Марина Дариовна	78 280 165,96 Р	248	724 816,35 Р
6	Мурашкин Сергей Евгеньевич	45 423 205,66 Р	14	4 129 382,33 Р
7	Цулауф Герман Яковлевич	43 084 756,55 Р	131	598 399,40 Р
8	Писаров Сергей Леонидович	36 545 522,00 Р	9	5 220 788,86 Р
9	предприниматель Волков Эдуард	24 988 275,40 Р	24	1 469 898,55 Р
10	ЩЁКИНА СВЕТЛАНА СЕРГЕЕВНА	24 364 905,67 Р	17	1 624 327,04 Р
11	Савин Сергей Борисович	23 358 385,26 Р	3	11 679 192,63 Р
12	ИП Васьков Николай	19 759 371,26 Р	27	823 307,14 Р
	Индивидуальный предприниматель			
13	Агагусейнов Магаммад Агарахим оглы	19 119 342,65 Р	35	1 062 185,70 Р
14	Лапаревич Наталья Александровна	11 578 316,75 Р	8	1 929 719,46 Р
15	Общий итог	326 502 247,16 Р	516	1 943 465,76 Р

Рисунок 4.3. Пример результата среза

Для построения диаграммы необходимо сделать срез куба. Он позволит на его основе сделать диаграмму. Пример диаграммы показан на рисунке 4.4.



Рисунок 4.4. Пример диаграммы

Для составления динамических показателей следует воспользоваться настройками PowerPivot и PowerView. Для использования PowerView следует предварительно сформировать запрос в PowerPivot.

Ниже показан пример отчёта, выполненного в PowerPivot, на основе затрат университетов с 2013 года по 1 триместр 2016 года. В нижней таблице указывается, на какой уровень бюджета производились закупки и какие суммы

потрачены. В правой части указываются затраты всех университетов, фрагментированные по годам (рисунок 4.5)

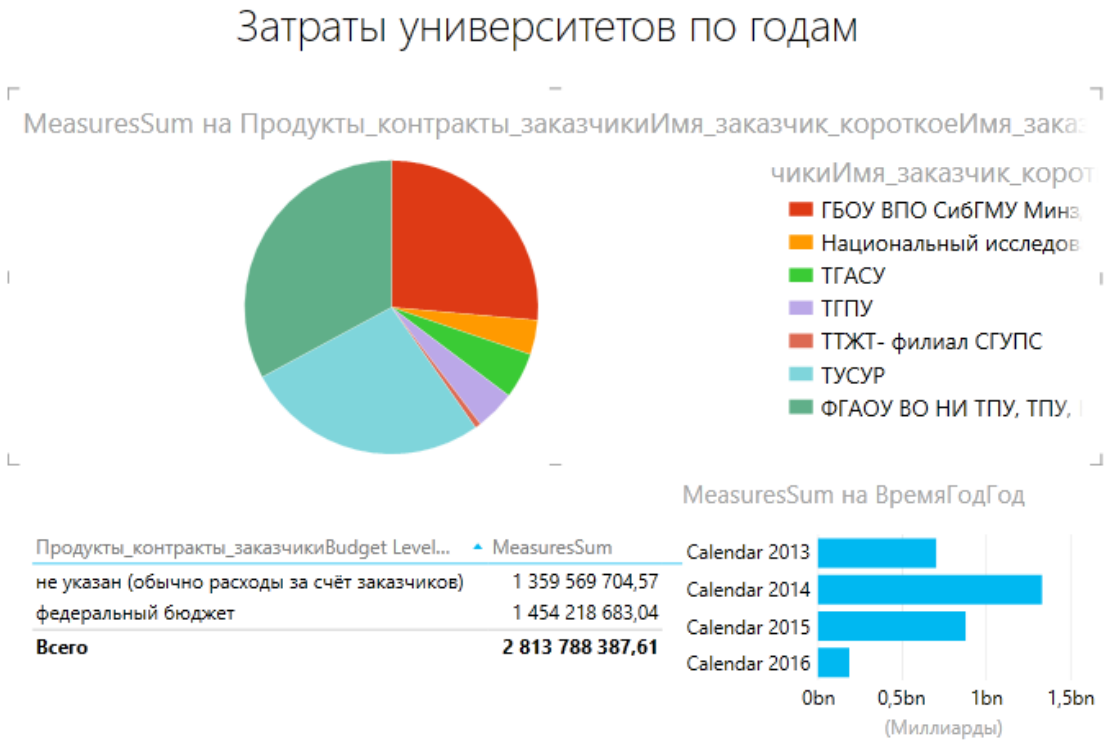


Рисунок 4.5. Пример диаграммы в надстройке PowerView

На диаграмме возможно выделить долю данных по интересующему университету и году (рисунок 4.6)

MeasuresSum на Продукты_контракты_заказчикиИмя_заказчик_короткоеИмя_заказчик_короткое

Категория	Доля (%)
ГБОУ ВПО СибГМУ Минздрав	~20%
Национальный исследовательский университет	~5%
ТГАСУ	~5%
ТГПУ	~3%
ТТЖТ- филиал СГУПС	~1%
ТУСУР	~25%
ФГАОУ ВО НИ ТПУ, ТПУ	~41%

MeasuresSum на ВремяГодГод

ВремяГодГод	MeasuresSum (Миллиарды)
Calendar 2013	~0.5
Calendar 2014	~1.2
Calendar 2015	~0.8
Calendar 2016	~0.3

Для использования возможностей карты необходимо при выборе
ммы нажать на кнопку Карты.

Ниже показан пример просмотра поставщиков на карте:

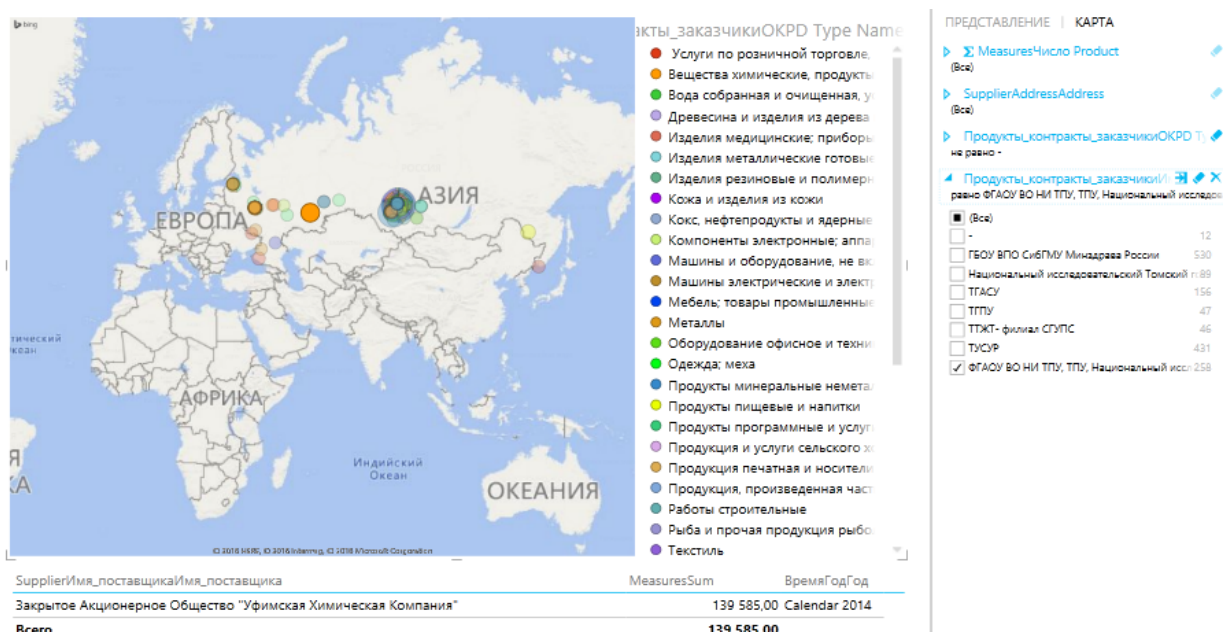


Рисунок 4.7. Пример отображения данных в надстройке PowerMap

В показанном окне отображаются заказы НИ ТПУ по городам России. Цвет обозначает тип заказа. В таблице ниже указана компания-поставщик, стоимость закупаемых объектов и год закупки. Справа находится область фильтров, в котором, можно выбрать отображение поставщиков, выполняющих заказы конкретного университета.

Привязка на карте выполняется по адресам. Заказы СибГМУ в пределах Томска за 2015 год представлены на рисунке 4.8

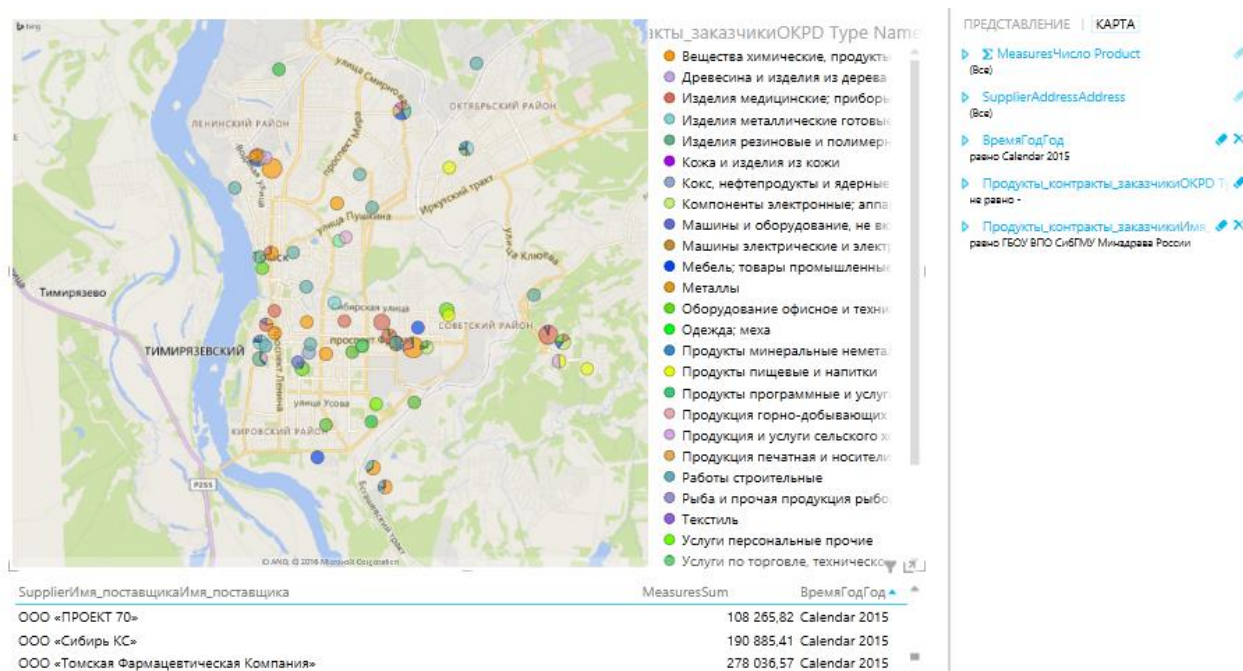


Рисунок 4.8. Пример масштабирования карты

Помимо привязки по адресам, возможно указать также регион заказчика. Регион указывает на головной офис фирмы поставщика, т.к. берётся, исходя из данных КПП кода. Так, банки, оказывающие услуги заказчикам Томской области, принадлежат главным образом регионам Москва и Санкт-Петербург.

Возможно также анализировать продукты во времени, используя точечную диаграмму (рисунок 4.9), либо вставляя параметр Год вертикальные или горизонтальные множители.

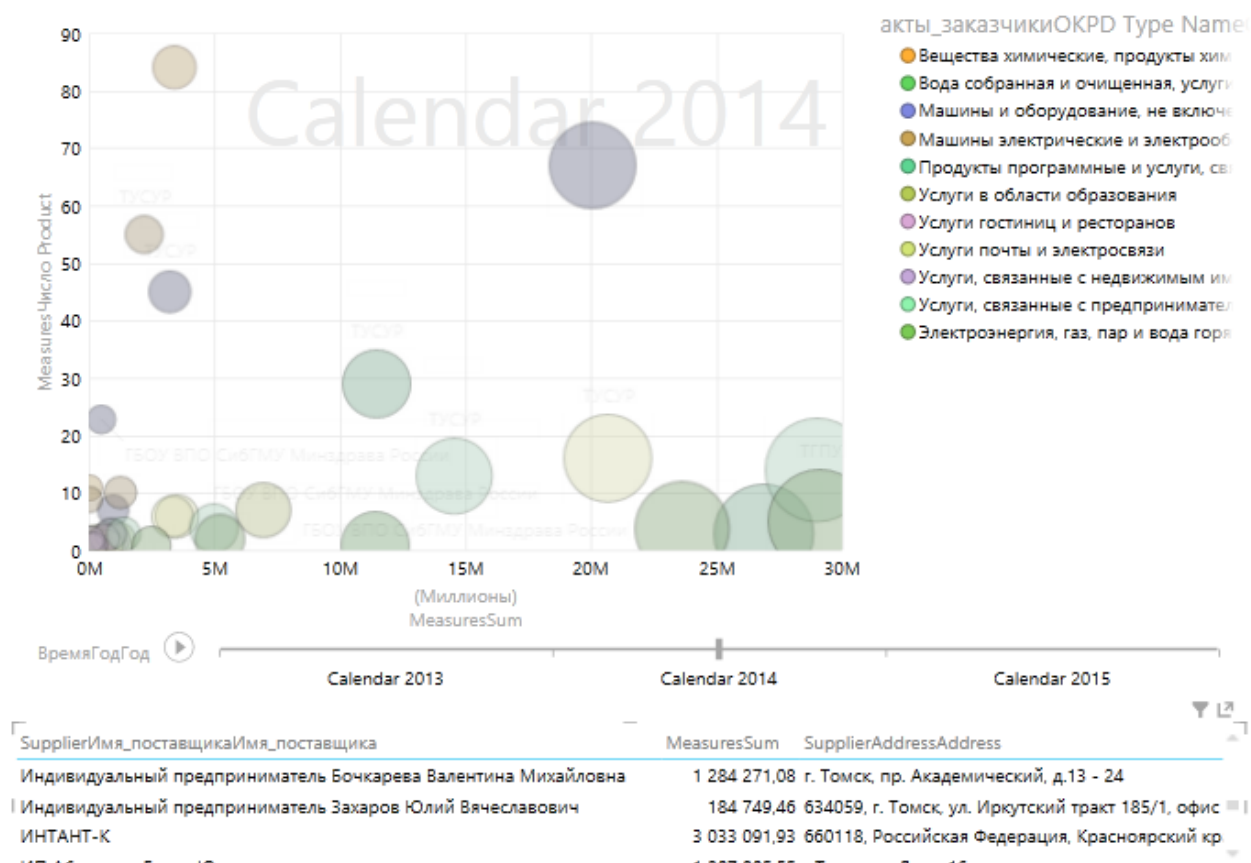


Рисунок 4.9. Пример использования точечной диаграммы

Здесь по вертикали указано, сколько заказано продуктов университетом, по горизонтали установлена мера стоимости заказа. Цветом обозначены типы ОКПД, а осью прокрутки являются года. При нажатии на интересующий объект снизу появится информация о поставщике и сумме заказа.

4.2 Примеры результатов анализа

Разработанный куб предоставляет широкие возможности для анализа данных портала госзакупок. Далее показаны примеры анализа функционирования институтов контрактной системы в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд Томской области, регулирующийся Законом №44-ФЗ. Примеры составляет лишь малую часть результатов, которые можно получить при использовании программного продукта.

На официальном сайте по Томской области с 2013 года по апрель 2016 года всего было заключено 56406 контрактов на сумму около 71,9 млрд рублей. Всего за время функционирования портала было размещено 56608 контрактов на сумму ФЗ 82,6 млрд рублей, действующих согласно закону №44-ФЗ.

Сведения о размещённых контрактах по способу определения поставщика показаны в таблице 4.1.

Таблица 4.1 – Сведения о размещенных контрактах в разбивке по процедурам определения поставщика

	Количество контрактов, шт				Прирост количества контрактов относительно прошлого года, шт		Прирост количества контрактов относительно прошлого года, %	
	2013	2014	2015	2016	2014	2015	2014	2015
Способ размещения								
Закупка единственного поставщика (подрядчика, исполнителя) у	3523	4329	12000	5472	806	7671	23%	177%
Электронный аукцион	930	11174	8905	2252	10244	-2269	1102%	-20%
Запрос котировок	540	2936	2002	382	2396	-934	444%	-32%
Открытый конкурс	40	473	83	18	433	-390	1083%	-82%
Запрос предложений	1	176	41	10	175	-135	17500%	-77%
Конкурс ограниченным участием с		206	2	1	206	-204	0%	-99%
Закрытый аукцион		1	6	5	1	5	0%	500%
Способ определения поставщика установлен в соответствии со статьёй 111 ФЗ				8	0	0	0%	0%
Закрытый конкурс			3		0	3	0%	0%
Закрытый конкурс с ограниченным участием		1	1		1	0	0%	0%
Общий итог	5034	19296	23043	8148	14262	3747	283%	19%

Самый большой прирост контрактов был в 2014 году. В 2015 году в Томске произошёл прирост закупок у единственного поставщика (+177%), остальные способы закупок, кроме закрытого аукциона, уменьшили свои показатели. В первом триместре 2016 года также чаще других осуществляются закупки у единственного поставщика. Примечательно, что при таком способе закупки не проводится конкурса, поставщик выбирается самостоятельно.

По стоимости способы определения поставщика в 2015 году распределились следующим образом:

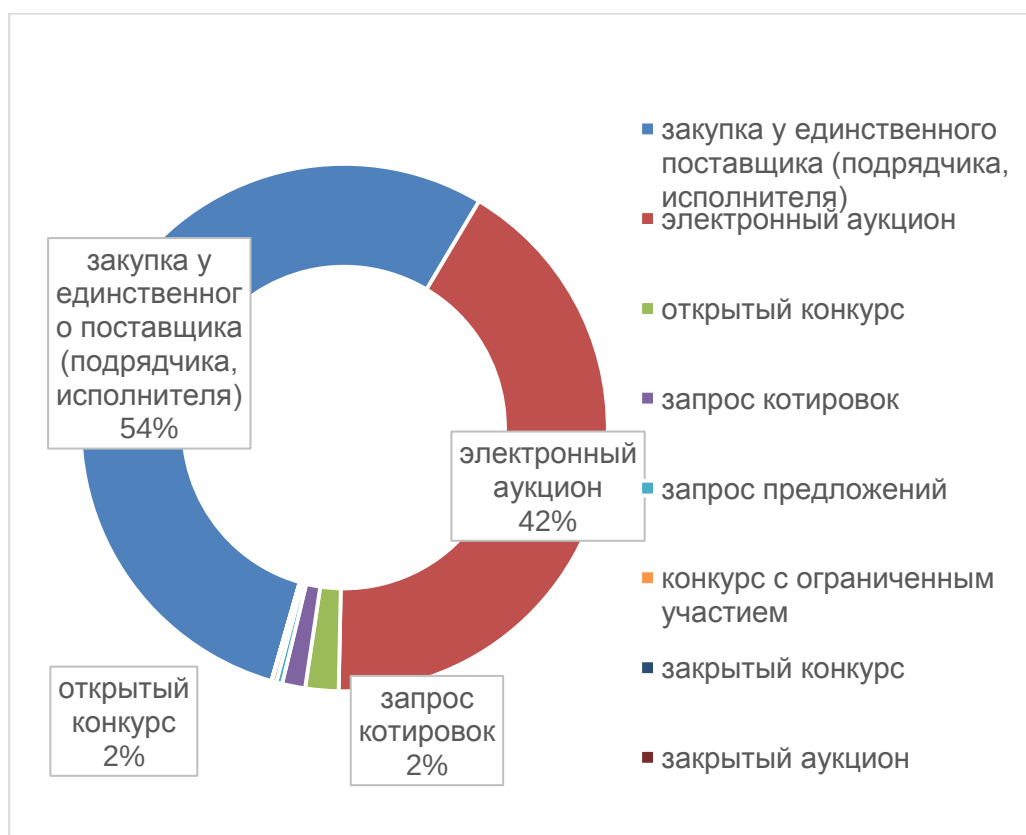


Рисунок 4.10. Сведения о заключённых контрактах за 2015 год относительно цены

Преобладающим способом определения поставщика является закупка у единственного поставщика, процент которого составляет 54% от общей суммы затрат на 2015 год. На втором месте находится электронный аукцион с 42%. Данные показатели отличаются от показателей по России. Согласно Докладу о результатах мониторинга применения Федерального закона №44-ФЗ за 2015 на

первом месте в России находится Электронный аукцион с 51,8%, закупка у единственного поставщика идёт следом с 21,7%.

Ниже представлен топ 10 крупнейших закупок у единственного поставщика в Томской области по итогам 2015 года:

Таблица 4.2 – Топ 10 закупок у единственного поставщика за 2015 год

Название фирмы заказчика	Название фирмы поставщика	Название объекта закупки	Общая сумма объекта закупки
Департамент финансов Томской области	Публичное акционерное общество "Совкомбанк"	Оказание финансовых услуг по предоставлению кредитов	615 578 082 Р
Департамент финансов Томской области	Публичное акционерное общество "Совкомбанк"	Оказание финансовых услуг по предоставлению кредитов	769 578 082 Р
Акционерное общество "сибирский химический комбинат"	Открытое акционерное общество "Свердловский научно-исследовательский институт химического машиностроения"	Разработка рабочей конструкторской документации, изготовление, поставка, проведение шеф-монтажа выпарной установки для переработки трапных вод и установки цементированья зданий 4А и 5 МФР	671 152 991 Р
Акционерное общество "сибирский химический комбинат"	Открытое акционерное общество "Свердловский научно-исследовательский институт химического машиностроения"	Разработка рабочей конструкторской документации, изготовление, поставка, проведение шеф-монтажа комплекса оборудования обращения с ТРО зданий 4А и 5	356 324 660 Р
Управление по распоряжению муниципальным имуществом и землей администрации Верхнекетского района	Общество с ограниченной ответственностью "Соцсфера Томской домостроительной компании"	Приобретение нежилого здания, 4-этажный (в том числе подземных -1), общая площадь 3 992,4 кв.м, адрес (местонахождение) объекта: Томская область, Верхнекетский район, р.п. Белый Яр, ул. Рабочая, 5а, предназначенное для размещения дошкольного образовательного учреждения на 220 мест	227 057 123 Р

Название фирмы заказчика	Название фирмы поставщика	Название объекта закупки	Общая сумма объекта закупки
Департамент капитального строительства администрации Города Томска	Томскремстрой-проект	Работы по капитальному ремонту МАОУ «Средняя общеобразовательная школа № 36» по адресу: МО «Город Томск», ул. Иркутский тракт, 122/1	181 267 291 Р
Управление дорожной деятельности, благоустройства и транспорта администрации Города Томска	ООО "Горсети"	услуги по организации освещения улиц и праздничного освещения МО «Город Томск» в 2016 - 2018 годах	400 984 700 Р
Управление дорожной деятельности, благоустройства и транспорта администрации Города Томска	Унитарное муниципальное предприятие «Спецавтохозяйство г. Томска»	Выполнение работ по комплексному содержанию автомобильных дорог, расположенных на территории муниципального образования «Город Томск», в 2016 г.	393 680 909 Р
Управление дорожной деятельности, благоустройства и транспорта администрации Города Томска	Унитарное муниципальное предприятие «Спецавтохозяйство г. Томска»	работы по ремонту улично-дорожной сети муниципального образования "Город Томск".	293 708 936 Р
Администрация Города Томска	Открытое акционерное общество 'Томская домостроительная компания'	участие в долевом строительстве многоквартирного(ых) дома(ов) в Городе Томске	327 370 050 Р

Самыми крупными неконкурентными закупками за 2015 год являются оказание финансовых услуг, закупки оборудования, закупки, связанные с приобретением недвижимости, работы по ремонту дорог и строительству.

Ниже приведена статистика по заключённым контрактам по месяцам с 2013 года по 1 триместр 2016 года.

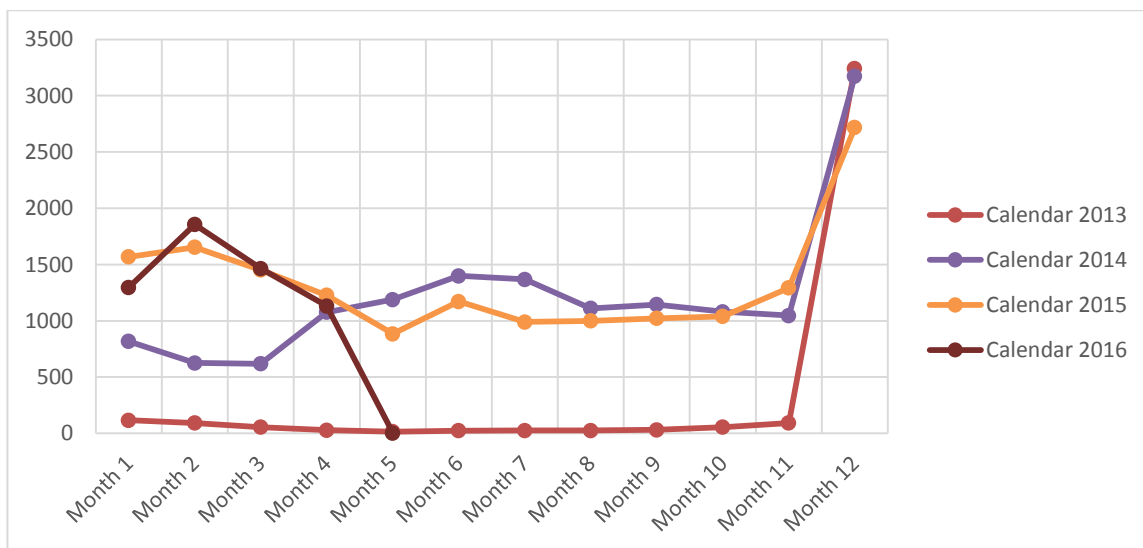


Рисунок 4.11. Объём ежемесячно заключённых контрактов в разбивке по годам относительно количества контрактов

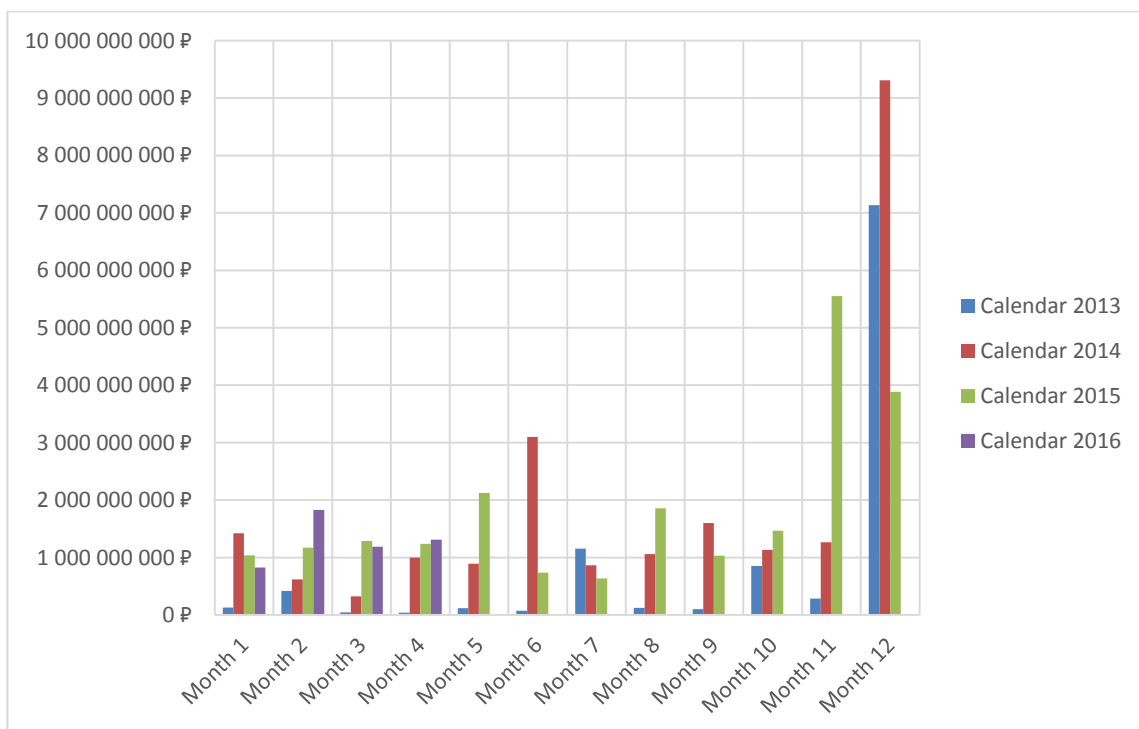


Рисунок 4.12. Объём ежемесячно заключённых контрактов в разбивке по годам относительно цены

Сильный скачок в декабре 2014г объясняется покупкой Сибирским химическим комбинатом комплекса производственных линий на 5,67 млрд рублей (рисунок 4.12)

Судя по графику, основной скачок закупок, регулируемых №44-ФЗ, произошёл в 2014г. Число контрактов, заключённых в 2015г, лишь незначительно превышает количество контрактов за 2014г.

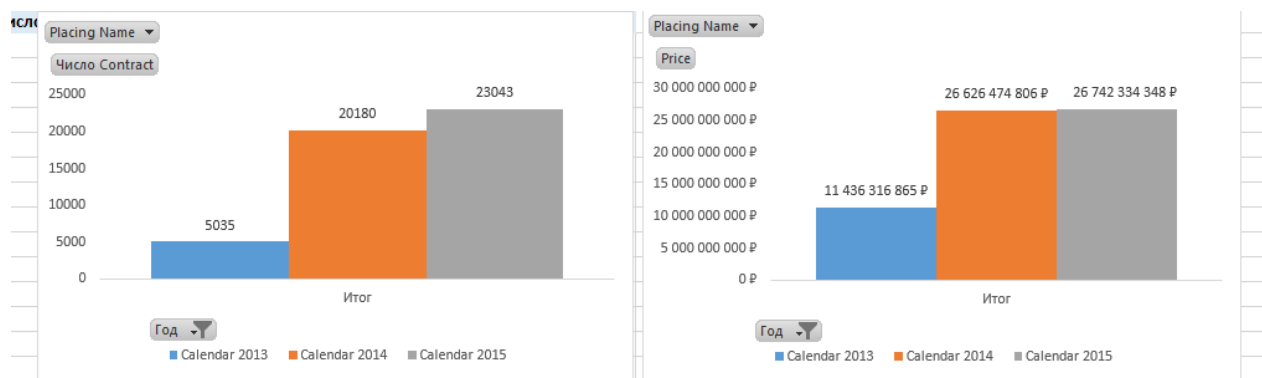


Рисунок 4.13. Сравнение количества и общей цены контрактов по годам

Анализ, отображённый на рисунке 4.13, показывает, что в 2015г по сравнению с 2014г, возросло число контрактов, но сумма примерно равна прошлогодней. Следовательно, средняя цена контракта уменьшилась (1,32 млн рублей в 2014 году и 1,16 млн рублей в 2015 году), что вполне может утверждать сокращение заказчиками сроков исполнения контрактов в целях снижения финансовых рисков поставщиков.

Произошёл рост объёма закупок у субъектов малого предпринимательства и социально ориентированных некоммерческих организаций. В 2015 году субъекты малого предпринимательства (СМП) и социально ориентированные некоммерческие организации (СОНКО) заключили 12833 контракта общим объёмом порядка 7,98 млрд рублей, что на 18,09% и 13,67% превышает показатели 2014 года. Таблица, показывающая количество поставщиков, заключённых ими контрактов и суммы контрактов у СМП и СОНКО, представлена ниже:

Таблица 4.3 - Показатели СМП и СОНКО за 2013-2015гг

Год	Число поставщиков	Число заключённых контрактов	Общая сумма контракта
2013	290	1679	1 626 769 839,21 Р
2014	1273	10867	7 017 579 657,76 Р

Год	Число поставщиков	Число заключённых контрактов	Общая сумма контракта
2015	1295	12833	7 977 205 425,50 Р
Общий итог	1697	25379	16 621 554 922,47 Р

Общий объем заключенных в 2015 году контрактов в разрезе номенклатуры ОКПД и ОКПД 2 в соответствии с рейтингом оснований их заключения в сравнении с аналогичными показателями 2014 и первого триместра 2016 года представлен на диаграммах 4.14 и 4.15.



Рисунок 4.14. Топ 10 типов закупок в разрезе номенклатуры ОКПД



Рисунок 4.15. Топ 10 типов закупок в разрезе номенклатуры ОКПД2

Самыми большими затратами за 2015 стали затраты на строительные работы (затраты по ОКПД и ОКПД 2 составили 6 млрд рублей). Маленькие затраты по ОКПД на 2016 год (около 673 тыс. рублей) связаны с постепенным переходом на ОКПД 2. Ранее Росстандарт ввел ОКПД2 в действие с 1 января 2014 года с правом досрочного применения в правоотношениях, возникших с 1 января 2014 года. Для перехода на классификаторы был установлен переходный период до 1 января 2016 года [http://ppt.ru/news/134780]. С 1 января 2017 года применения ОКПД 2 станет обязательным.

Услуги по финансовому посредничеству занимают второе место по стоимости контрактов, на которые затрачено 4,6 млрд рублей. Самыми популярными объектами закупки за 2015 год являются:

- «Оказание финансовых услуг по предоставлению кредитов», где заказчиком является Департамент финансов Томской области. Стоимость контрактов составляет 3,8 млрд рублей.

- «Предоставление кредитных ресурсов муниципальному образованию «Город Томск»», где заказчиком является «Департамент финансов администрации Города Томска». Стоимость контрактов составляет 758 млн рублей.
- «Оказание услуг по финансовой аренде (лизингу) медицинского оборудования» больницами Томской области. Стоимость контрактов составляет 10,4 млн рублей.

Примечательно, что по всей России на втором месте по стоимости контрактов занимают закупки медицинских изделий и фармацевтических препаратов. В томской области эти закупки находятся на 3 месте с общей стоимостью объектов закупок порядка 3,5 млрд рублей (значение складывается из типов ОКПД «Вещества химические» и «Изделия медицинские»).

Анализ заказов объектов закупки в период с 2013 года по апрель 2016 года, руководствующихся Федеральным законом №44-ФЗ, в разрезе по месяцам, представлен на диаграмме 4.16.



Рисунок 4.16. Объем ежемесячно заказываемых объектов закупки в разбивке по типам ОКПД

Таким образом, в Томской области чаще других заказывают медицинские препараты и изделия (категория Вещества химические). По количеству объектов закупки строительные изделия находятся на 5 месте.

Пик заключения контрактов приходится на декабрь. Это обусловлено тем, что компаниям необходимо выполнить закрытие бюджета, выделенного на текущий год.

Ниже рассмотрена статистика по образовательным учреждениям.

Согласно portalу госзакупок, количество дошкольных учреждений в Томской области - 148. Количество школ при этом составляет 279. При этом расходуются деньги бюджетных учреждений, федеральный бюджет и собственные средства. Расходы собственных средств за 2015 год составляют 4,88% у детских садов и 10,22% у школ.

Школы в основном тратятся на энергию, строительство, оказание услуг питания и продукты.

Детские сады также, только добавляются предпринимательские услуги - под этим подразумеваются услуги по уборке, обслуживанию зданий и охране.

Университеты больше закупают мебель, вычислительную технику, услуги, связанные с научной деятельностью. Затраты на закупку электроэнергии в университетах меньше в 11 раз, чем у школьных учреждений и составляют 34,6 млн рублей за 2015 год. СибГМУ в больших объёмах закупает медицинские изделия и лекарства.

Ниже рассмотрена статистика по поставщикам.

Доля поставщиков с иных регионов упала по сравнению с 2014 годом на 4,23 %. Однако сумма заключения контрактов с поставщиками Томской области выросла на 3,84%

Таблица 4.4 – Сравнение поставщиков Томской области с поставщиками с других регионов

	Количество поставщиков, шт			Общая сумма контрактов, млрд рублей		
Года	2013	2014	2015	2013	2014	2015
Томская область	1410	3267	2761	10,26	15,59	16,68
Все регионы	1636	4752	4279	11,46	26,61	26,73
% заказов у поставщиков Томской области от других регионов	86,19%	68,75%	64,52%	89,39%	58,57%	62,41%

Ниже показано, в каких регионах находятся поставщики, заключившие контракты с Томской областью. Информация о регионе берётся из кода постановки на учёт, который привязан к месту регистрации предпринимателя.

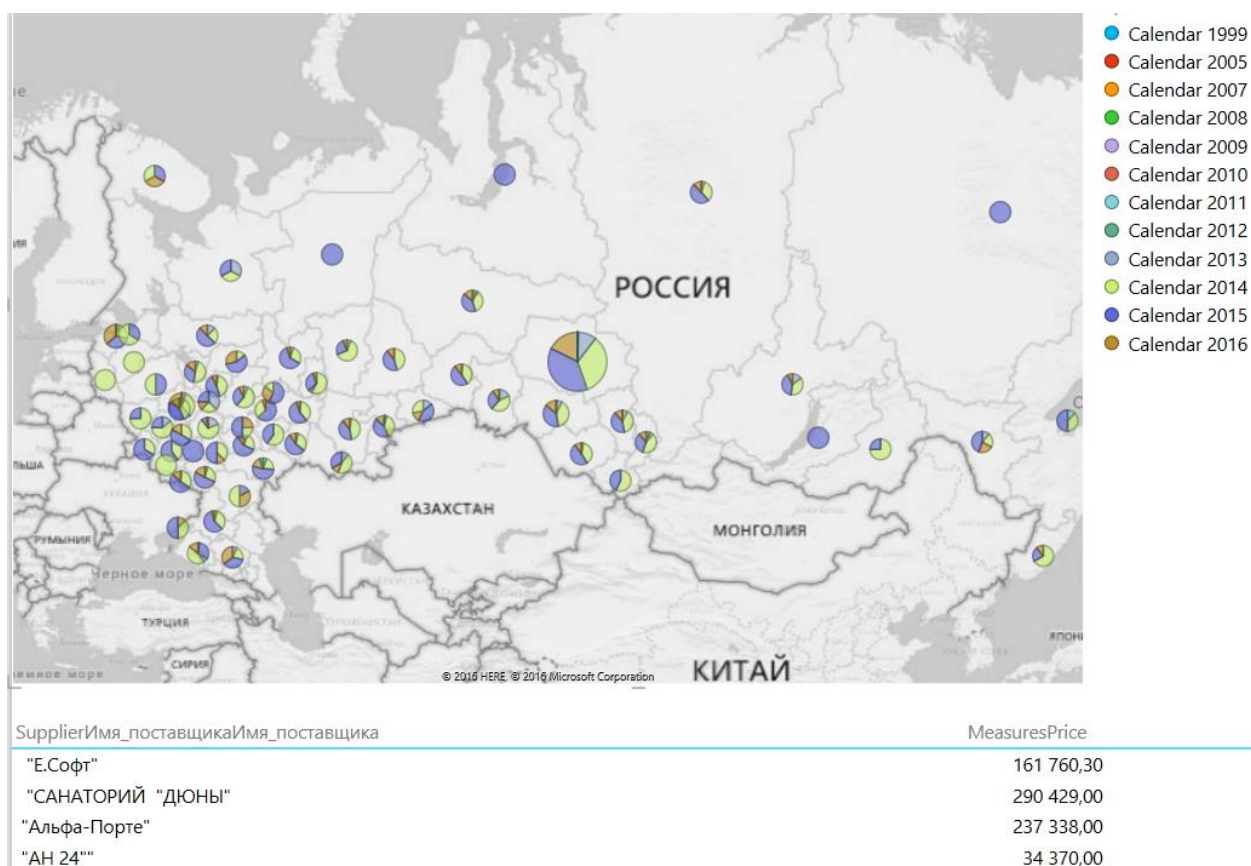


Рисунок 4.17. Отображение поставщиков по регионам

Объём круга зависит от количества поставщиков в регионе. Круг разделён на секторы по годам, в которые поставщики региона участвовали в закупках Томской области. В таблице снизу отображается информация о названии фирмы поставщика и общей сумме контрактов. Таблица фильтруется по году и региону России.

4.3 Нарушения

Разработанный продукт позволяет выявлять различные виды нарушений, включая нарушения, связанные со сроками выполнения, с закупками у единственного поставщика затруднение поиска заказа, накрутка стоимости, закупку товаров для собственных нужд. Ниже рассмотрены примеры поиска нарушений.

4.3.1 Нарушения со сроками

В измерениях куба присутствуют поля *difStartEndDay*, которые содержат разницу между датами конца и начала исполнения условий контракта, которые

следует использовать при выявлении подобных нарушений. Примером нахождения нарушения является случай, когда заказчик разместил заказ на создание и развёртывание ПО со сроком исполнения всего 6 дней, в последних числах декабря (таблица 4.5). Однако в среднем разработка, установка ПО и обучение сотрудников составляет несколько месяцев.

Таблица 4.5 – Пример неправильного заполнения сроков

Наименование услуги	Дата заключения контракта	Дата начала исполнения	Дата окончания исполнения	Стоимость услуги
Оказание услуг по разработке программного приложения переноса данных для нужд Администрации Томского района	25.12. 2015	25.12. 2015	31.12.2015	18 тыс. рублей

Ссылка на контракт: <http://zakupki.gov.ru/epz/contract/contractCard/common-info.html?reestrNumber=3701404452215000093>

4.3.2 Анализ применения латиницы в русских словах

Иногда одна или несколько букв в названии тендера пишется латиницей, что затрудняет поиск заказа на портале. Для выявления подобных случаев нарушений был написана программа на скриптовом языке Python. Запрос программы показан в приложении Г.

В программе используется кодировка символов ASCII, т.е. выполняется поиск латинских символов, идущих за кириллицей. Для нахождения рассматриваемого нарушения используется среда SQL Server Management Studio 2012. Необходимо подключиться к базе данных и применить скрипт нахождения нарушений.

Результаты выявления нарушений находятся в таблице *LatynAnalysis*. Часть кода программы, выявляющая нарушения, представлена в Приложении Г

Стоит заметить, что не все найденные результаты являются нарушениями, т.к. имеют место опечатки, частично иностранные названия и т.д. Однако встречаются и стабильные ошибки у определённых заказчиков.

Помимо замены букв алфавита имеют место случаи намеренного указания буквы «О» вместо цифры «0» и наоборот.

Под подозрением в нарушениях находятся 432 контракта.

Ниже приведены названия фирм-заказчиков, которые используют подмену букв чаще других:

Таблица 4.6 – Подозрительные фирмы-заказчики

Название фирмы заказчика	Количество контрактов	Дата заключения последнего контракта	Позиция в строке	Символ лат-кир	Невалидное слово	Общая сумма подозрительных контрактов
ОГКУ "Томсклес"	197	18.04.2016 г.	8	от	моторное	2 771 544 Р
ТУСУР	130	9.02.2016 г.	1	То	Точка	3 248 635 Р
ФГБУ "Сибирский федеральный научный клинический центр Федерального медико-биологического агентства"	85	19.04.2016 г.	34	оС	SvetoCopy	9 294 663 Р
НИИ кардиологии	55	22.04.2016 г.	26	ХХ	XXXL	15 510 526 Р
ГБОУ ВПО СибГМУ Минздрава России	43	29.12.2015 г.	43	Ом	500мл	862 228 Р
ОГБУЗ "Чаинская РБ"	37	18.03.2016 г.	3	рн	Чернила	128 287 Р
ОГБУЗ "Лоскутовская РП"	27	11.01.2016 г.	48	te	Microvette	2 128 965 Р
ОГБУЗ "ОДБ"	27	28.12.2015 г.	8	ук	фруктовое	307 288 Р
ОГБУЗ "Томский областной центр по профилактике и борьбе со СПИД и другими инфекционными заболеваниями"	26	19.04.2016 г.	101	uC	TruCal	8 301 571 Р
ОГБУЗ "Каргасокская РБ"	20	25.09.2014 г.	20	te	Monovette	1 028 767 Р
Администрация Ленинского района Города Томска	18	23.11.2015 г.	110	пФ	inФОРМАТ	73 851 Р
ОГБУЗ "Асиновская РБ"	16	26.10.2015 г.	95	РА	РАМГ-1	359 151 Р
ФБУЗ "Центр гигиены и эпидемиологии в Томской области"	16	30.03.2015 г.	82	АС	АРТАСА	186 177 Р
ОГБУЗ "Бакчарская РБ"	14	12.04.2016 г.	61	gM	IgM	160 398 Р
ОГБУЗ "Кривошеинская РБ"	14	18.08.2015 г.	1	ТО	ТОКСОПЛАЗМОЗ	182 561 Р

Название фирмы заказчика	Количество контрактов	Дата заключения последнего контракта	Позиция в строке	Символ лат-кир	Невалидное слово	Общая сумма подозрительных контрактов
ОГБУЗ "Верхнекетская РБ"	13	24.02.2016 г.	44	ал	альфа-амилазы	146 565 Р
ОГБУЗ "Томский региональный центр крови", ОГБУЗ "ТРЦК"	13	25.01.2016 г.	1	Ка	Карточки	4 355 336 Р
ФКУ "ЦУКС ГУ МЧС России по Томской области"	11	25.05.2015 г.	19	пФ	ТекстМарк. InФармат	19 643 Р

Больницы в основном меняют символы в местах названий химических элементов либо при указании длины в см.

ТУСУР допускает ошибки в таких словах как «Рабочая чаСтота». В данном словосочетании указаны английские р и с.

ОГКУ "Томсклес" совершает нарушения чаще других, и, в отличие от других заказчиков, названии объекта закупки почти всегда написано на русском языке без обильного использования иностранных слов, например: «Предоставление услуг подвижной радиотелефонной связи стандарта GSM 900/1800 (Свойства : /.)»

ЗАКЛЮЧЕНИЕ

В ходе данной работы были выполнены следующие задачи:

- собрана и проанализирована научно-техническая информация по тематике исследования;
- разработана информационная модель данных о государственных закупках;
- проанализированы типовые мошеннические схемы;
- проведен анализ существующих решений;
- разработана архитектура информационно-аналитической системы;
- реализована база данных;
- построены витрины данных для проведения многомерного анализа;
- спроектированы и реализованы ETL процессы;
- проведен анализ полученных данных
- найдены мошенники на основе анализа применения латиницы

Во время работы над магистерской диссертацией были написаны статьи «Использование технологий Business Intelligence для анализа данных в сфере государственных закупок», которая была размещена в материалах международной научно-практической конференции «Технологии Microsoft в теории и практике программирования» весной 2015 года и «Извлечение и анализ данных с портала государственных закупок», которая была размещена в материалах международной научно-практической конференции «Молодежь и современные информационные технологии».

Проведено комплексное описание и анализ финансово-экономических аспектов выполненной работы. Рассмотрены вопросы выполнения требований к безопасности и гигиене труда, к промышленной безопасности, охране окружающей среды и ресурсосбережения.

СПИСОК ПУБЛИКАЦИЙ

1. Ибраева Н.С.; Сергеев Д.А.; Кудинов А.В.; Использование технологий Business Intelligence для анализа данных в сфере государственных закупок // Технологии Microsoft в теории и практике программирования: сборник трудов XII Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых (Томск, 25–26 марта 2015 г.) / Томский политехнический университет. – Томск: Изд-во Томского политехнического университета, 2015. – 251 с.
2. Д.А. Ибраева, А.В. Кудинов; Извлечение и анализ данных с портала государственных закупок // Молодежь и современные информационные технологии : сборник трудов XIII Международной научно-практической конференции студентов, аспирантов и молодых ученых (г. Томск, 9-13 ноября 2015 г) / Томский политехнический университет. – Томск: Изд-во Томского политехнического университета, 2015. – С. 84-85.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Выгрузка условий конкурсов госзакупок с ООС zakupki.gov.ru / [Электронный ресурс] / Хабрахабр. URL: <http://habrahabr.ru/post/253201/>, свободный. Загл. с экрана. – Яз. рус. Дата обращения: 19.03.2016 г.
2. Способы закупок по 223-ФЗ и требования к документации [Электронный ресурс] / Электронная торговая площадка для торгов по закупкам 223-ФЗ и коммерческих закупок – OTC-tender. URL: http://www.otc.ru/tender/sposoby_zakupki_trebovaniya_k_documentacii.aspx, свободный. – Загл. с экрана. – Яз. рус. Дата обращения: 19.03.2016 г.
3. Как проводится контроль в сфере закупок [Электронный ресурс] / Разъяснения в редакции, действующей с 21.01.2014 (ПК «Система главбух»); М.В. Чемерисов. Электрон. Дан. — 2014.— 3 с. URL: http://www.yamaledu.org/activity/economics_and_finance/state_order/arhiv-gosudarstvennyy-zakaz/8226-kak-provoditsya-kontrol-v-sfere-zakupok-mv-chemerisov.html, свободный. Загл. с экрана. – Яз. рус. Дата обращения: 13.05.2016 г.
4. Российская практика проведения государственных закупок [Электронный ресурс] / Совершенствование процесса проведения государственных закупок. Рекомендации и материалы, подготовленные Рабочей группой промышленных компаний (Industrial Companies Working Group); Мухтаров Э.В., Жуков В.Н. Электрон. Дан. – М., 2012. – 64 с. URL: http://www.iblfrussia.org/upload/iblock/239/Improvement_of_public_procurement_procedures_ru.pdf, свободный. Загл. с экрана. – Яз. рус. Дата обращения: 19.03.2016 г.
5. Разъяснения по процедуре выгрузки сведений об опубликованных документах на FTP-сервер Общероссийского официального сайта [Электронный ресурс] / Документация портала государственных закупок – Руководства пользователя. URL: <http://zakupki.gov.ru/epz/main/public/>

- download/downloadDocument.html?id=793, свободный. – Загл. с экрана. – Яз. рус. Дата обращения: 13.05.2016 г.
6. Пособие для выявления сомнительных закупок [Электронный ресурс] / Редакционная группа проекта «За честные закупки». Электрон. Дан. 2 изд. – 8 с. URL: https://zachestnyezakupki.onf.ru/files/Ostap_book_2.pdf, свободный. – Загл. с экрана. – Яз. рус. Дата обращения: 13.05.2016 г.
 7. Информация о системе «Закупки 360» [Электронный ресурс] / Startpack. – URL: <https://startpack.ru/application/zakupki360>, свободный. – Загл. с экрана. – Яз. рус. Дата обращения: 19.03.2016.
 8. ПО «ГосЗакупки» [Электронный ресурс] / URL: <http://www.soft-zakupki.ru/>, свободный – Загл. с экрана. – Яз. рус. Дата обращения: 19.03.2016 г
 9. О проекте [Электронный ресурс] / ГосЗатраты. URL: <https://clearspending.ru/page/about/> свободный – Загл. с экрана. – Яз. рус. Дата обращения: 19.03.2016 г
 10. Решения по построению корпоративных хранилищ данных [Электронный ресурс] / IBA Rus. URL: <http://ibarus.ru/solutions/dwh/>, свободный. Загл. с экрана. – Яз. рус. Дата обращения: 15.05.2016 г.
 11. Лекция 2: Многомерное представление данных. Схемы звезда и снежинка [Электронный ресурс] / Интуит. URL: <http://www.intuit.ru/studies/courses/1145/214/lecture/5508?page=2>, свободный. – Яз. рус. Дата обращения: 19.05.2016 г.
 12. David Loshin. ETL (Extract, Transform, Load) // Business Intelligence. — 2nd. — Morgan Kaufmann, 2012. — 400 p.
 13. Тарифы закупки 360 [Электронный ресурс] – URL: http://zakupki360.ru/#?slide=cta_tariffs, свободный. – Загл. с экрана. – Яз. рус. Дата обращения: 2.05.2016.
 14. Белов С.В. Безопасность жизнедеятельности и защита окружающей среды: учебник для вузов. – М.: Изд- во Юрайт, 2013. – 671с.

15. СанПиН 2.2.4.1294-03 «Гигиенические требования к аэроионному составу воздуха производственных и общественных помещений».
16. СанПиН 2.2.2/2.4.1340 – 03 «Гигиенические требования к персональным электронно–вычислительным машинам и организации работы». – М.: Госкомсанэпиднадзор, 2003.
17. ГОСТ 12.1.030 -81. Защитное заземление, зануление.
18. СНиП 21 – 01 – 97. Пожарная безопасность зданий и сооружений. М.: Гострой России, 1997. – с.12.
19. Федеральный закон Российской Федерации от 22 июля 2008 г. N 123-ФЗ "Технический регламент о требованиях пожарной безопасности".
20. Гусельников М.Э., Методические указания по разработке раздела «Производственная и экологическая безопасность» выпускной квалификационной работы для студентов всех форм обучения.
21. Гусельников М.Э., Методические указания по разработке раздела «Производственная и экологическая безопасность» выпускной квалификационной работы для студентов всех форм обучения.

ПРИЛОЖЕНИЕ А

A.1 Government procurement portal

Portal of the state purchases of the Russian Federation [1] - The official website of the Russian Federation in the network "Internet" for posting information about placing orders for goods, works, services (hereinafter - the official website) is designed to provide a free and gratuitous access to complete and accurate information about the contract system in the procurement and the procurement of goods, works and services, certain types of legal entities, as well as for the formation, processing and storage of such information.

The goal of the portal is to ensure the transparency of transactions, operational and mass informing about the upcoming auction, raising the competitiveness of suppliers, efficient spending of state funds, and, therefore, improving the quality of implementation of state and municipal needs.

The basic concepts in the field of public procurement are:

- Public procurement is the process of getting goods, works and services by public institutions.
- Contract is a public, municipal agreement, a civil contract.
- Customer is represented by a state body, governing agency for non-budgetary funds administration, local authority, budget organization and another recipient of the federal budget of the Russian Federation and local budgets.
- Supplier - is legal entity (organization, enterprise, institution) or natural person supplying goods, products or services to customers.
- Purchase can be defined as a set of actions starting by placing the notice on the procurement implementation for state needs.
- The product in the context of master's thesis - a certain set of goods / products / services of the same content. In one contract may be ordered a few products.

The bidding organizer can be one of the customers or bodies authorized for bidding and tenders.

The public procurement process in Russia are regulated by the Federal law No. 44-FZ "On the System of Public Procurement Contracts for Products, Works, Services for State and Municipal Needs" as of April 4, 2013. The majority of the 44-FZ provisions will come into effect on January 1, 2014, on the same date Federal Law No. 94-FZ "On Placement of Orders to Supply Goods, Carry out Works and Render Services for State and Municipal Needs" ceased to be in force. 223-FZ "On Purchase of Goods, Works and Services by Special Types of Legal Entities" as of July 18, 2011, regulates purchases, performed by special types of legal entities. In this master's thesis purchases, subjecting to the Law No. 44-FZ, is analyzed.

Here are basic paradigms of the Federal Law No. 44-FZ:

- ensuring publicity and transparency in procurement;
- prevention (or minimization) of corruption and other abuse;
- improving the quality of ensuring the state and municipal needs;
- implementation of the system approach in the field of procurement.

The purpose of this mechanism is to improve the control over spending the budgetary funds, ensure the efficiency and transparency of public procurement, and also develop the necessary analytical system in decision making process.

The main stakeholders of public contracts are divided into four groups:

- customer;
- supplier;
- regulator;
- general public.

The Federal Law No. 44-FZ dated April 4, 2013 regulates the following aspects of the relationship in the public procurement:

- procurement planning for state and municipal needs;
- publication of the notification about state orders;

- accepting applications for participation;
- holding a tender;
- contract signing;
- contract execution.

The placing way of the tender (method of determining supplier) is chose during purchase planning and documents preparation. The article No. 24 (The Federal Law No. 44-FZ) elects the following placing ways:

- *Open tender.* The winner is the participant who offered the best conditions of the order. The bidders are evaluated on a number of criteria in addition to cost of the contract and the quality characteristics of goods, including the estimated business participant 's reputation, experience, terms of delivery of goods, etc.
- *Competition with limited participation.* This placing way is used for innovation-oriented, technical and technological sophisticated purchases. Thus, suppliers must confirm qualification level to take a part in the tender.
- *The two-stage tender.* The winner is the party to have taken part in both stages of the competition (including prequalification at the first stage in case of additional requirements for the participants), and offered the best conditions of contract execution at the second stage of the competition.
- *Closed tender.* The same as the open tender, but customer communicates procurement information by sending invitations for participation in closed methods of determining suppliers.
- *Closed tender with limited participation.* The same as the open tender with limited participation, but customer communicates procurement information by the sending invitations for participation.

- *Closed two-stage tender.* The same as open two-stage tender, but customer communicates procurement information by sending invitations for participation.
- *Auction in electronic form.* Holding a public auction is carried out only in electronic form. The winner is the participant who has offered the lowest contract price.
- *Closed auction.* The same as open auction, but customer communicates procurement information by sending invitations for participation.
- *Request for proposals.* Information about purchased goods is reported to general public by publication of the notification. The winner is the participant of the purchase, who sends the final proposal that best meets the requirements.
- *Request for quotation.* Information about need for goods is reported to general public by publication of the notification. The winner is the participant who has offered the lowest contract price.
- *Purchase from a single supplier.* It is not a competitive procedure in contrast to previous procedures. The Federal Law No. 44-FZ determines the list of cases (reasons) for purchase from a single customer.

All-Russia official website on government procurement is an unified information resource with access to placed information. The site is located at <http://zakupki.gov.ru>.

All-Russia official website allows customizing the search and filtering of competitions by parameters. Search on the site is not always correct. The site is not stable, sometimes there are disconnects. This can affect to extraction process of information process for continuous replenishment of the data base. The site allows downloading the search results in CSV format, but with a limit of 500 rows.



ИНФОРМАЦИЯ О КОНТРАКТЕ № 1701731112716000015

ISS - подписка на события контракта ?

Размещено 25.05.2016 16:17 (MSK+3 (UTC+6) Новосибирск)
По местному времени заказчика

ИНФОРМАЦИЯ О
КОНТРАКТЕ

ИНФОРМАЦИЯ О ПЛАТЕЖАХ
И ОБЪЕКТЕ ЗАКУПКИ

ИНФОРМАЦИЯ ОБ ИСПОЛНЕНИИ
(РАСТОРЖЕНИИ) КОНТРАКТА

ДОКУМЕНТЫ

ЖУРНАЛ
СОБЫТИЙ

ОБЩАЯ ИНФОРМАЦИЯ

Реестровый номер контракта	17017311127 16 000015
Статус контракта	Исполнение
Номер извещения об осуществлении закупки	0865100000116000003
Способ определения поставщика (подрядчика, исполнителя)	Электронный аукцион
Дата подведения результатов определения поставщика (подрядчика, исполнителя)	16.02.2016
Дата размещения (по местному времени)	25.05.2016 16:17 (МСК+3)
Реквизиты документа, подтверждающего основание заключения контракта	Протокол подведения итогов электронного аукциона № 0865100000116000003-3 от 16.02.2016

ИНФОРМАЦИЯ ОБ ИЗМЕНЕНИИ КОНТРАКТА

Основание для изменения	изменение суммы исполнения этапа
-------------------------	----------------------------------

Figure A.1. Example of a solicitation on the government procurement website

In addition to posting information on the site <http://zakupki.gov.ru/>, portal uploads information to FTP-server which contains a structured version of the site information. To FTP-server load regularly updated information about ongoing competitions.

Addresses of FTP servers are divided by a federal laws: Federal Law №223; Federal Law №94 and Federal Law №44. The directories which contain information about the Federal Law №94 and №44, located at <ftp.zakupki.gov.ru> and have the following structure:

- *94fz* is directory containing the data of public uploading in accordance with №94 FZ (other directories - #44FZ);
- *fcs_nsi* is directory containing the data of regulatory and reference information;

- *fcs_regions* is directory containing the full details of the regional unloading published on All-Russia official website;
- *fcs_banks* is directory containing the unloading data of bank guarantees, published on the All-Russia official website;
- *fcs_rules* is directory containing information of specification
- *fcs_fas* is directory containing the unloading data information about suspicious suppliers and audits, the All-Russia official website;
- *fcs discussion* directory containing unloading of discussions.

The server is stored archived files in XML (extensible Markup Language), which contain information about the published documents (notices, minutes, contracts, etc.). In one archive can be no more than 3000 files.

The main data source is the FTP-server of public procurement portal.

All documents are uploaded on FTP servers in two ways: full regional uploading and automatic uploading by the rules [5]. Files on the server are divided into directories according to a relevant region. Full regional upload includes the following types of documents: published notifications, published changes of notifications, published protocols, published information about contracts, published changes of contracts, published information about performing or cancellation of contracts.

Files in the unloading dividing of the directories, the relevant region. Each region directory contains such directories: SketchPlans, PlanGraphs, Notifications, PurchaseDocs, Protocols, Contracts, AddInfo, CustomerReports, RegulationRules, RequestQuotation, AuditResult. Such each directory contains additional directories currMonth and prevMonth.

Figure A.2 shows the structure of the “44fz” folder.

<pre> – < Region Name>; – notifications; – currMonth; </pre>

```

        - prevMonth;
    - protocols;
        - currMonth;
        - prevMonth;
    - contracts;
        - currMonth;
        - prevMonth;
    - <...>
- < Region Name>;

```

Figure A.2. The structure of the “44fz” folder on the FTP server

A list of documents published for the previous day is daily uploaded on the server. A list of documents published for the previous calendar month is monthly uploaded.

If at the moment of uploading formation for the past period there have been no published document, the XML-file is unloaded empty. After the monthly upload directories with daily uploadings for the past month are cleared.

A.2 Irregularities in government contracts

Existing irregularities schemes in government contracts are discussed below. There are irregularities carried out by a trading platform and outside. For example, there are document forgery and sharing of confidential information to competitors. Fraudulent transactions, that capture the procurement system, can be tracked by analyzing information from the website, while the blackmail and sharing of confidential information can be revealed entirely different methods. The most common schemes suitable for detection by data mining methods are presented below.

Purchases with excess consumption properties or overestimated purchase price. This is usually a purchase of very expensive goods or services from proved suppliers. There are unwarranted budget spending, government spendthrift and the possible limiting of competition when concluding such contracts. This case can be

found in the published notice. The initial cost must be checked for overstatement regarding the market. Then the expert will be required to view the documentation for any special requirements for materials, services, goods. The initial price of the contract can be unsubstantiated or substantiated incorrectly.

Open tender announcement instead of the auction. This scheme is used to include quality indicators and get rid of unnecessary vendors using specially prepared tender documentation, which is suitable for a competition, but in fact they buy the products which require the auction. This leads to limitation of competition and increases the cost of the contract.

Contract signing not with the winner bidder. The customer concludes the contract not with the participant, who offered a lower price, but with those who offered more money. As major violations there are overestimation of the total contract price, the restriction of competition in favor of loyal companies, wasteful budgetary spending.

Splitting Large Requirements. Contracting or requiring activity personnel may be splitting requirements into small purchase orders to avoid the scrutiny required for larger cost value contracts. Splitting the requirement may waste funds by losing the economic advantage of volume purchasing. The essence of the scheme is a reduction of the minimum period of notification for limiting competition in favor of the companies which have been aware of the upcoming tender.

Unrealistic deadlines. Customer specifies unrealistic terms of the contract in the specifications or in design and estimate documentation. Thus, the specified amount of work within this period can be executed only by one contractor who has already done a part of the order, having started execution previously before the notification was published. This scheme is based on the relationships between representative of government agency and supplier.

Complication of the search in databases. Sometimes one or more letters in the title of the tender is written in Latin or typos are made in keywords. Suppliers cannot find the order inserting the correctly typed option. The orders, that have a title with

extra spaces, dashes or underscores, are included in the same category. Also, each service or product has its own All-Russian Product Classification code, which is important for companies-suppliers.

Rigged Specifications. The requesting organization tailors developing specifications to meet the qualifications of one particular company, supplier, or product. This applies mostly to procurement of medicines.

Most of the fraudulent schemes based on the collusion of procurement participants. This may be a collusion between the contractors or between procurement organizers and contractors.

To detect some fraudulent schemes the analysis of the following purchasing documents is needed, such as customer requirements specification, design and estimate documentation etc. This information is quite heterogeneous and stored as files of different formats. Syntactic analysis of these documents will increase the information extraction time and could not lead to the expected result, because recognition and text mining is a separate direction in the artificial intelligence.

To reveal suspicious transactions using data mining techniques the following scheme are best suited: luxury products and overestimated purchase price, contract signing not with the winner bidder, open tender announcement instead of the auction, splitting large requirements, unrealistic deadlines, complication of the search in databases and rigged Specifications.

Fraudsters often operate for some well-known schemes. Knowing some scheme, heuristic analysis can be applied to recognize these behaviors and prevent them. But by analyzing a large number of procurement, much more effective would be to use an automated data mining techniques such as OLAP and Data Mining.

ПРИЛОЖЕНИЕ Б

Схема базы данных



ПРИЛОЖЕНИЕ В

Полное содержание атрибутов мер и измерений

Информация о контракте:

- Номер изменения
- Способ определения поставщиков (примеры: электронный аукцион, запрос котировок)
- Основание заключения контракта с единственным поставщиком (заполняется, если проводилась закупка у единственного поставщика)
- Номер лота
- Цена контракта
- Валюта
- Дата подведения результатов определения поставщика
- Дата заключения контракта
- Дата публикации контракта
- Дата начала исполнения контракта
- Дата окончания исполнения контракта
- Уровень бюджета (примеры: федеральный бюджет, местный бюджет, бюджет Фонда социального страхования Российской Федерации)
- Вид бюджетных средств (примеры: Бюджет ЗАТО Северск, Областной бюджет, Бюджет муниципального образования "Город Томск")
- Гиперссылка на опубликованные сведения о контракте
- Текущее состояние контракта (примеры: исполнение, исполнение завершено)

Заказчик:

- Полное наименование
- Сокращённое наименование
- Организационно-правовая форма организации в ОКОПФ
- ИНН
- КПП
- Код по ОКПО

Объект закупки:

- Классификация по ОКПД
- Классификация по ОКПД 2
- Тип ОКПД
- Тип ОКПД 2
- Единицы измерения
- Цена за единицу в валюте контракта
- Количество поставляемого товара, объёма выполняемых работ, оказываемых услуг
- Стоимость поставляемого товара, выполняемых работ, оказываемых услуг в валюте контракта
- Наименование товара, работ, услуг

Поставщик

- Тип участника (примеры: физическое лицо, юридическое лицо, индивидуальный предприниматель)
- ИНН
- КПП
- Полное наименование
- Сокращённое наименование
- Фирменное наименование поставщика

- Код по ОКПО
- ОКТМО (примеры: город Тюмень, город Новосибирск)
- Код региона (примеры: Томская область, Москва)
- Адрес
- Организационно-правовая форма организации в ОКОПФ (примеры: Общество с ограниченной ответственностью, Федеральное казенное предприятие)
- Статус (примеры: организация инвалидов, предприятие уголовно-исполнительной системы, субъект малого предпринимательства, социально-ориентированная некоммерческая организация)

Время

- Год
- Месяц
- Дата

Меры измерения:

- Количество заказчиков
- Количество поставщиков
- Количество контрактов
- Цена контракта (суммирование)
- Количество продуктов
- Цена за 1 единицу продукта (суммирование)
- Количество товара в одном продукте
- Стоимость поставляемого товара
- Цена контракта (вычисление среднего значения)
- Цена за 1 единицу продукта (вычисление среднего значения)

ПРИЛОЖЕНИЕ Г

Примеры кода программного продукта

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<ns2:export xmlns="http://zakupki.gov.ru/oos/types/1"
xmlns:ns2="http://zakupki.gov.ru/oos/export/1"
xmlns:ns3="http://zakupki.gov.ru/oos/printform/1">
  <ns2:contract schemeVersion="4.5">
    <id>17494083</id>
    <regNum>0165100003014000083</regNum>
    <signDate>2014-10-20+07:00</signDate>
    <versionNumber>0</versionNumber>
    <customer>
      <regNum>01651000030</regNum>
      <fullName>Прокуратура Томской области</fullName>
      <inn>7021018851</inn>
      <kpp>702101001</kpp>
    </customer>
    <price>161095.00</price>
    <currency>
      <code>RUB</code>
      <name>Российский рубль</name>
    </currency>
    <products>
      <product>
        <sid>64816148</sid>
      </product>
    </products>
    <suppliers>
      <supplier>
        <postAddress>131000, г.Москва, Варшавское шоссе, 37</postAddress>
        <contactPhone>7-3822-514269</contactPhone>
        <contactFax>7-3822-514269</contactFax>
      </supplier>
    </suppliers>
    <href>http://zakupki.gov.ru/epz/contract/contractCard/common-
info.html?reestrNumber=0165100003014000083</href>
    <currentContractStage>E</currentContractStage>
  </ns2:contract>
</ns2:export>
```

Рисунок Г.1. Пример XML файла

```
from os import listdir
import xml.etree.ElementTree as ET
import os
import pyodbc
from ftplib import FTP

ftp = FTP('ftp.zakupki.gov.ru', 'free', 'free')
ftp.cwd('/fcs_regions/Tomskaja_obl/contracts/')
files=ftp.nlst() # список архивов

all_files = []
cnxn = pyodbc.connect('DRIVER={SQL Server};SERVER=localhost;DATABASE=zakupki;UID=sa;
PWD=4321')
cursor = cnxn.cursor()
for row in cursor.execute(""" SELECT [archive] FROM [zakupki].[dbo].[allArchive]
```



```

    """):
    d = row.xml_file      # обработанные архивы
    all_files.append(d)
    cnxn.commit()

```

Рисунок Г.2. Пример кода скрипта загрузки и распаковки файлов с FTP

```

from os import listdir
import xml.etree.ElementTree as ET
import os
import pyodbc
import datetime
cnxn = pyodbc.connect('DRIVER={SQL Server};SERVER=localhost;DATABASE=contractsTomsk;
UID=sa; PWD=4321')
cursor = cnxn.cursor()
cnxn.commit()
i=0
count = 0
files = listdir(".")
xml_files = filter(lambda x: x.endswith('.xml'), files)
for xml_file in xml_files:      # 1 цикл - 1 файл
    count = count + 1
    open_file = open(xml_file,encoding = 'UTF-8' )
    using_file=open_file.read()
    doc = ET.fromstring(using_file) # s текст xml
    ns = {'oss': 'http://zakupki.gov.ru/oos/types/1'} # задаёт пространство имён
    tcus_inn = '-'
    tcus_kpp = '-'
    tid = '-' # заполнение значений тире на случай отсутствия данных
    tpublishDate = '1800-01-01'
    <...>
    for child in doc[0]:      #1lvl      - 1 ур вложенности
        if '{http://zakupki.gov.ru/oos/types/1}id' == child.tag: # поиск по тегам
            tid = child.text
        elif 'regNum' in child.tag:
            tregNum = child.text
        <...>
        for child1 in child:      #2lvl
            elif 'executionDate' in child.tag and 'month' in child1.tag:
                tmonth = child1.text
            elif 'executionDate' in child.tag and 'year' in child1.tag:
                tyear = child1.text
                tendDate = datetime.datetime(int(tyear), int(tmonth), int(tday))
                #print ('    tendDate:',tendDate)
            elif 'executionPeriod' in child.tag and 'startDate' in child1.tag:
                sDate = child1.text[0:10]
                tstartDate = datetime.datetime.strptime(sDate, '%Y-%m-%d')
                #print('    tstartDate: ',tstartDate)
                <...>
                tsupInn = '-' # чтобы не попадало значение с уровня выше (иначе в 4 лвл
появляется лишнее кпп)
                tsupKpp = '-'
                for child2 in child1:      #3lvl
                    if 'lotNumber' in child2.tag:
                        tlotNumber = child2.text
                    <...>
                    for child3 in child2:      #4lvl
                        if 'lotNumber' in child3.tag:
                            tlotNumber = child3.text

```

```

        <..>
        cursor.execute("""IF NOT EXISTS(SELECT * FROM [contractsTomsk].[dbo].[Contract]
        WHERE [fileVersion] = ? and [cusInn] = ? and [cusKpp] = ? <..>)
        BEGIN
        INSERT INTO Contract ([fileVersion],[cusInn],[cusKpp],<..>)
        VALUES (?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?) --
23 атрибута
        END""", (tfileVersion,tcusInn,tcusKpp, <..>))
        cnxn.commit()
        #os.remove(xml_file)      # удаляет файл после прочтения
        if (count%500)==0:
            print ('Конец обработки файла №', count)
            print (datetime.datetime.now().time())
        cnxn.close()

```

Рисунок Г.3. Часть кода скрипта, парсящего контракты

```

DECLARE
@productName nvarchar(2000),
@код int,
<..>
create table LatynAnalys (id nvarchar(20), sid nvarchar(15), код int, код1 int,
символ char, символ1 char, позиция int, строка nvarchar(2000))
SET @CURSOR = CURSOR SCROLL FOR
SELECT id, sid, productName FROM [contractsTomsk].[dbo].[Product]
OPEN @CURSOR
FETCH NEXT FROM @CURSOR INTO @id, @sid, @productName
WHILE @@FETCH_STATUS = 0
BEGIN
    SET @position = 1
    --select @productName
    WHILE @position <= DATALENGTH(@productName)
    BEGIN
        SET @код= ASCII(SUBSTRING(@productName, @position, 1))
        set @код1 = ASCII(SUBSTRING(@productName, @position+1, 1))
        IF (@код not between 182 and 255) and @код !=168 <..>
        BEGIN
            if (@код1 between 188 and 255) or @код1 =168 or @код1 =184
            BEGIN
                INSERT INTO LatynAnalys
                SELECT @id as id, @sid as sid, @код as код, @код1 as код1,
                CHAR(ASCII(SUBSTRING(@productName, @position, 1))) as
символ, CHAR(ASCII(SUBSTRING(@productName, @position+1, 1))) as символ1,
                @position, @productName as productName
            END -- 2if
        END -- 1 if
        SET @position = @position + 1
    END
    FETCH NEXT FROM @CURSOR INTO @id, @sid, @productName
END --1 while
CLOSE @CURSOR
SELECT * FROM LatynAnalys -- в таблицу помещаются все последовательности
латинский символ/кириллица
GO -- конец

```

Рисунок Г.4. Часть кода программы, выявляющей замену кириллицу латиницей