Министерство образования и науки Российской Федерации

ФГАОУ ВО «Севастопольский государственный университет»

Институт информационных технологий и управления в технических системах

Кафедра Информационных систем

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

к курсовому проекту

по дисциплине «Технологии проектирования информационных систем»

Выполнил:

ст. гр. ИС/м-11о Криворучко Ю.Э.

Проверил:

проф. Доронина Ю.В.

Севастополь

2017

**СОДЕРЖАНИЕ**

[1 ПОСТАНОВКА ЗАДАЧИ 3](#_Toc479718406)

[1.1 Описание предметной области 3](#_Toc479718407)

[1.2 Описание входных и выходных данных ИС 5](#_Toc479718408)

[1.3 Описание требований к ИС 7](#_Toc479718409)

[1.4 Описание ограничений при создании ИС 9](#_Toc479718410)

[2 ОПИСАНИЕ КОМПЛЕКСА ТЕХНОЛОГИЙ ДЛЯ ПРОЕКТИРОВАНИЯ ИС 12](#_Toc479718411)

[2.1 Обоснование выбора модели жизненного цикла при проектировании ИС 12](#_Toc479718412)

[2.2 Описание вложенных уровней проектирования ИС 13](#_Toc479718413)

[2.3 Обоснование комплекса технологий при проектировании ИС 14](#_Toc479718414)

[2.4 Формализованное описание комплекса технологий при проектировании ИС 17](#_Toc479718415)

[3 ОЦЕНКА ЭФФЕКТИВНОСТИ ВЫБРАННОГО КОМПЛЕКСА ТЕХНОЛОГИЙ ПРИ ПРОЕКТИРОВАНИИ ИС 19](#_Toc479718416)

[3.1 Выбор критериев оценки комплекса технологий при проектировании ИС 19](#_Toc479718417)

[3.2 Расчет показателя эффективности оценки комплекса технологий 20](#_Toc479718418)

[4 ОЦЕНКА ВОЗМОЖНЫХ ЭКСПЛУАТАЦИОННЫХ ХАРАКТЕРИСТИК ИС ЗАДАЧЕ НИР 23](#_Toc479718419)

[4.1 Расчет информационных характеристик БД (требуемый объем жесткого диска при учете интенсивности приращения информационного фонда) 23](#_Toc479718420)

[БИБЛИОГРАФИЧЕСКИЙ СПИСОК 26](#_Toc479718421)

# **1 Постановка задачи**

# **1.1 Описание предметной области**

В рамках НИР выполняется оптимизация системы сжатия банковских данных, которая была разработана с выпускной квалификационной работе бакалавра. Система может быть применена в банковской сфере для сжатия банковских данных, например, типовых банковских договоров.

Расположение разрабатываемой системы в мегасистеме представлено на рисунке 1.1. Банковская сфера оперирует с большим количеством документом, в частности с договорами. Банк заключает договоры как с физическими и юридическими лицами на оказание услуг, так и договоры с налоговой службой для предоставления информации о банковских вкладах клиентов. В свою очередь налоговая служба может наложить арест на банковских вклад налогоплательщика, который уклоняется от уплаты налогов и передать информацию в таможенную службу для препятствования выезда за пределы страны. Налоговая служба также контролирует отчисления в пенсионный фонд и фонд социального страхования.

Информационная система банка включается в себя различные системы [4]:

* система ежедневных внутрибанковских операций (СВО);
* система коммуникаций с филиалами и иногородними отделениями;
* системы автоматизированного взаимодействия с клиентами «банк-клиент»;
* аналитические системы. Анализ всей деятельности банка и системы выбора оптимальных в данной ситуации решений;
* системы межбанковских расчетов;
* системы автоматизации работы банка на рынке ценных бумаг.

Система сжатия банковских данных ССБД является частью информационной системы банка. Таким образом, надсистемой для разрабатываемой системы является банк. Данные, которые являются входными для ССБД, поступают из хранилища данных, расположенного в надсистеме. Выходные данные из ССБД поступают в надсистему банк.

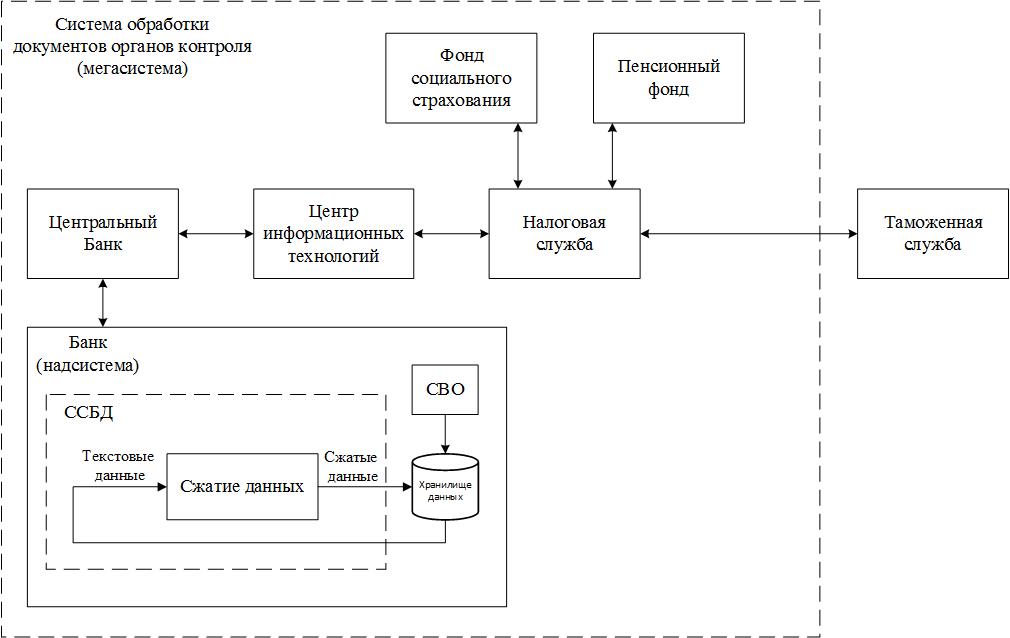


Рис. 1.1. Определение разрабатываемой системы в мегасистеме.

Рассмотрим преобразование данных в информационной системе с помощью диаграммы потока данных (рисунок 1.2). На вход системы поступают текстовые банковские данные на русском языке. Подробное описание входных и выходных данных приведено в пункте 1.2. Выполняется частотный анализ данных. В результате выполняется уточнение вероятностной модели текста, которая фиксируется в хранилище данных. В процессе сжатия исходный текст преобразуется, с помощью вероятностной модели в сжатые данные.

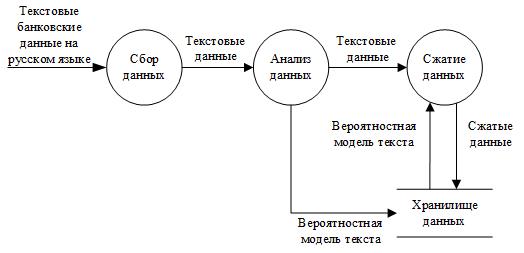


Рис. 1.2. Диаграмма потока данных в ССБД.

Выходными данными, разрабатываемой ССБД, являются преобразованные входные данные посредством сжатия.

# **1.2 Описание входных и выходных данных ИС**

В банковской сфере основным из документов преимущественно является договоры.  На рисунке 1.3 представлена классификация банковских документов. Данная классификация отражает лишь часть документов в банковской сфере. Банковские договоры можно разделить на следующие типы: договор купли-продажи валюты и ценных бумаг, договор банковского вклада, договор банковского счета, договор банковской гарантии, договоры займа и кредита. Каждый тип можно классифицировать по признакам, например, кредитные договоры можно разделить на договоры по форме заключения, по способу привлечений средств, по сроку заключения, по виду валюты. В среднем объем договора составляет 12кб – это примерно 800 символов.

Содержание договора зависит от того к какому типу он относится. Однако, можно отметить что в договоре содержатся условия заключения договора, реквизиты и адреса банка и клиента, а также наименование сторон, которые заключают договор. В договорах в банковской сфера с одной стороны обязательно выступает банк, а с другой стороны физическое либо юридическое лицо. Также договор может быть трехсторонним, например, договор банковской гарантии. В таком договоре сторонами выступают: гарант (банк, который обеспечивает гарантию), принципал (кто просит гарантии) и бенефициар (тот, кому гарант будет выплачивать денежные средства).

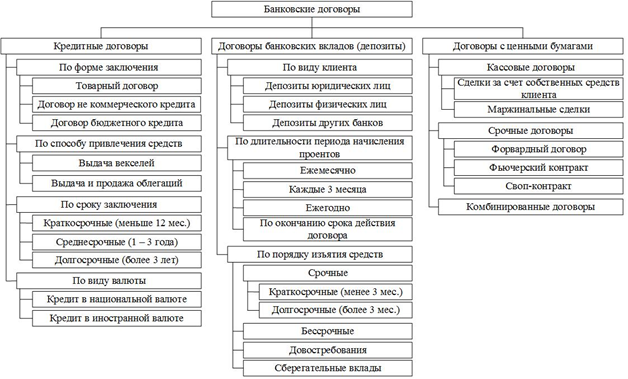


Рис. 1.3. Классификация банковских договоров.

Основываясь на проведенном предварительном анализе банковских текстов можно сказать, что в текстах банковской тематики наблюдаются следующие особенности: большое количество однокоренных слов, например, вкладчик, вклад, вкладчика или кредитор, кредит и так далее. Также в текстах присутствуют повторяющиеся устойчивые фразы такие как: именуемый в дальнейшем, расчетный счет, обязанности вкладчика. В каждом банковском документе указывается название банка, с которым заключен договор. Такое название может меняться в зависимости от регионального расположения банка: Симферопольский филиал АО КБ «Северный кредит», Московский филиал АО КБ «Северный кредит».

На рисунке 1.4 изображены входные и выходные данные для ССБД. Входные данные di являются документами в формате txt. Выходные данные d'i - документы в формате dat.

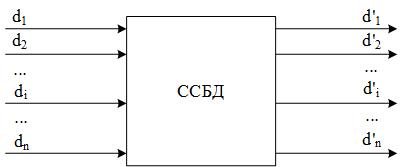


Рис. 1.4. Входные и выходные данные для ССБД.

Входными данными для ССБД являются банковские договоры расширением txt. Выходные данные получается посредством преобразования входных данных - сжатие данных. Сжатые данные представляют собой последовательность символов записанных в файл с расширением dat. По сравнению с исходными данными, сжатые данные занимают меньше места (памяти)  при хранении.

# **1.3 Описание требований к ИС**

ИС должна соответствовать следующим требованиям: требования к надежности данных G1, требования к доступу к данным G2, требования к безопасности системы G3, требования к параметрам системы G4.

Для того чтобы система соответствовала описанным требованиям необходимо выполнить оптимизацию критериев, описанных путем уточнения глобальной цели функционирования ИС – сжатие данных G0 (рисунок 1.5).

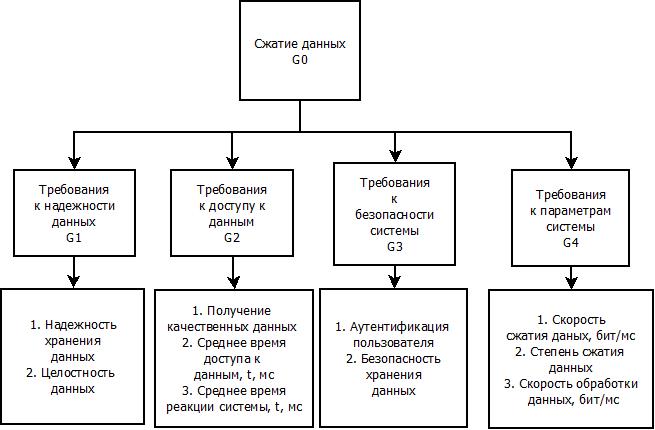


Рис. 1.5.Дерево требований к ССБД.

Требование к надежности данных G1 выполняется при реализации надежного хранения данных и целостности данных. Требование к доступу к данным G2 обеспечивается за счет реализации следующих критериев: получение качественных данных, среднее время доступа к данных, среднее время реакции системы. Требование к безопасности системы G3 реализуется при наличии аутентификации системы и безопасного хранения данных. Требования к параметрам системы G4 характеризуются скоростью сжатия данных, степенью сжатия данных, скоростью обработки данных.

# **1.4 Описание ограничений при создании ИС**

Рассмотрим разрабатываемую систему в рамках методологии функционального моделирования процессов (методология IDEF0). На рисунке 1.6 изображен главный процесс системы. Главный процесс системы – сжатие данных.

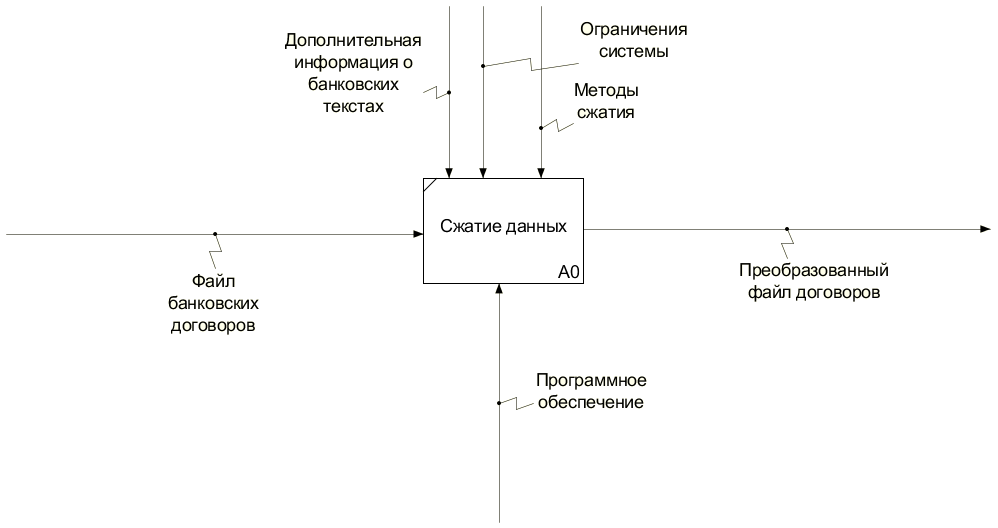


Рис. 1.6. Главный процесс ССБД (контекстная диаграмма).

На вход системы поступают текстовые банковские данные на русском и английском языках. На выходе системы – сжатые данные. Механизмом в разрабатываемой системе является программное обеспечение. Управляющей информацией является ограничения системы, дополнительная информация о банковских текстах, методы сжатия. Дополнительная информация о банковских текстах – это информация, полученная на этапе анализа банковских данных и представленная в виде таблицу. В эту таблицу входят одиночные символы, наиболее часто встречающиеся N-граммы и устойчивые словосочетания, а также сопоставленные с этими лексемами кодовые комбинации. В качестве метода сжатия в системе используется модифицированный словарный метод LZW. На рисунке 1.7 изображена декомпозиция главного процесса.

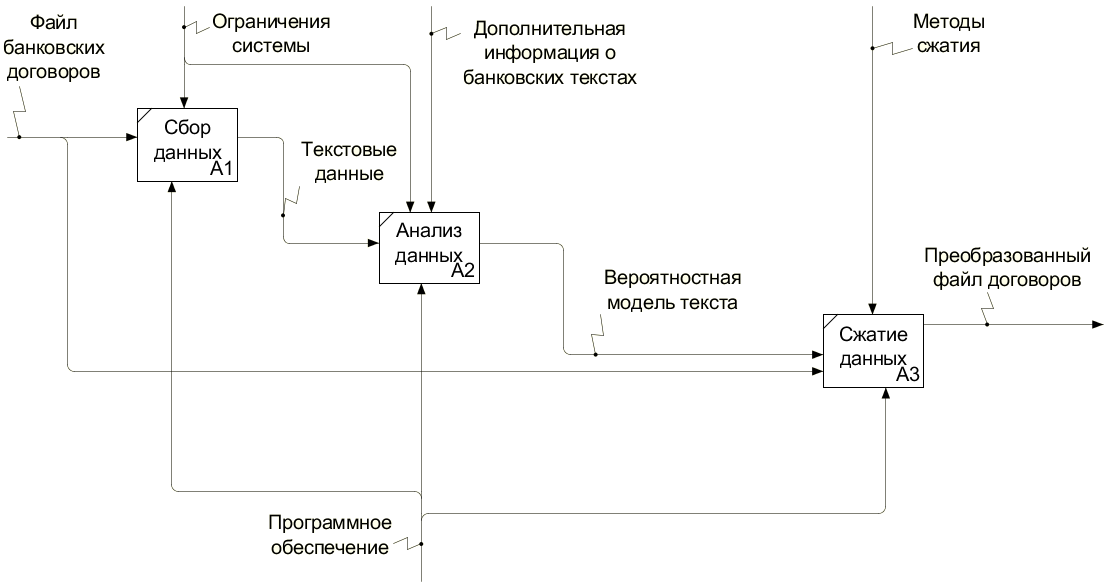


Рис. 1.7. Декомпозиция главного процесса.

В качестве ограничений системы можно выделить следующие:

* сжатие данных выполняется посредством модифицированного алгоритма LZW;
* таблица с дополнительной информации не может превышать 4096 строк;
* таблица с дополнительной информации включает в себя следующие лексемы: одиночные символы, N-граммы (N изменяется от 2 до 4) и устойчивые словосочетания, состоящие из двух либо трех слов;
* текст для сжатия должен содержать только символы русского и английского языков, также допускаются специальные символы;
* для сжатия используются текстовые файлы с расширением txt; у файлов после сжатия расширение dat;
* ограничение по объему сжимаемого файла зависит от производительности системы, на которой будет эксплуатироваться ИС.

# **2 Описание комплекса технологий для проектирования ИС**

# **2.1 Обоснование выбора модели жизненного цикла при проектировании ИС**

Для разрабатываемой системы выбрана разновидность поэтапной модели жизненного цикла, не инкрементная либо итерационная, так как при разработке системы в рамках НИР требования определяются на ранних этапах и не будут меняться в ходе проектирования и разработке. Также в инкрементной либо итерационной моделях разработка системы выполняется по частям. На каждом этапе выполняется полное прохождения стадий классического жизненного цикла от анализа требования до разработки и тестирования. Таким образов в конце каждого этапа выходит новая версия продукта с учетом изменившихся требований.

Жизненный цикл разрабатываемой ИС можно представить в виде каскадной модели [3]. При таком подходе последующая стадия начинается после полного завершения предыдущей. Наличие последовательности работ позволяет легко оценивать сроки завершения стадии и всей работы в целом. Однако, данная модель имеет недостаток - внесение изменений в завершенную стадию. Поэтому для разрабатываемой ИС выбрана модель жизненного цикла с возвратами (рисунок 2.1). Для данной модели необходимо ввести следующее ограничение: при создании ИС в рамках НИР на стадии разработки и тестирования не возвращаться к стадиям анализа предметной области, формирования требований и перепроектирование.

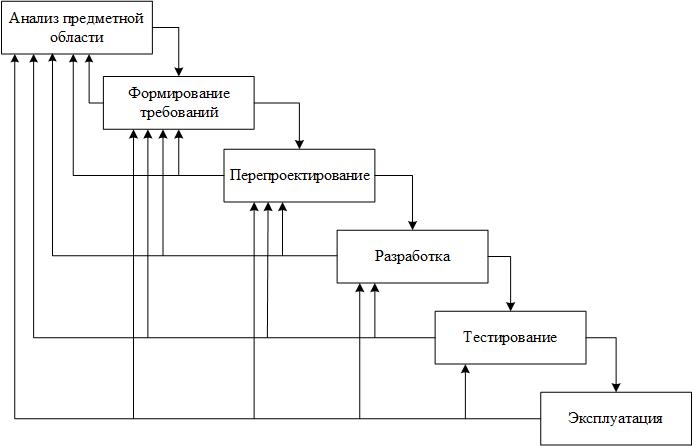


Рис. 2.1. Жизненный цикл ССБД.

Разрабатываемая ИС является усовершенствованной версией системы сжатия банковских данных, которая в отличие от предыдущей производит вероятностный анализ исходного файла данных, что позволяет улучшить степень сжатия данных. Поэтому классический этап в жизненного цикла ИС – проектирование изменен на перепроектирование.

# **2.2 Описание вложенных уровней проектирования ИС**

Диаграмма Венна, которая демонстрирует включение разрабатываемой системы в информационную систему банка. Информационная система – совокупность программно-технических средств, обеспечивающих сбор, обработку, накопление, хранение и дальнейшую передачу информационных ресурсов [2]. ССБД является составляющей информационной системы. Помимо ССБД информационная система банка включает в себя другие системы, которые были описаны в разделе 1.1. Используя все доступные составляющие. информационная система обеспечивает информационные потребности банка. Информационная система является частью структуры банка. Также в структуру банка включаются сотрудники банка. В свою очередь банк является частью бизнес-среды. Банк функционирует с другими предприятиями для обеспечения своих бизнес-целей.

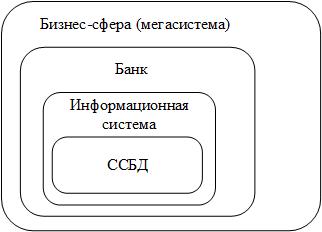


Рис. 2.2. Диаграмма Венна, демонстрирующая включение разрабатываемой системы в информационную систему банка.

# **2.3 Обоснование комплекса технологий при проектировании ИС**

Для построения схем используется графический редактор Dia. Программа бесплатная и кроссплатформенная, поэтому была выбрана среди аналогов. В данном продукте есть возможность описания новых объектов, экспорт и сохранение диаграмм в различных форматах, использование надстроек, которые позволяют автоматически создавать UML-схемы из программного кода, а также автоматическое преобразование UML-схем в программный код.

Для построения диаграммы DFD используется кроссплатформенная программа Ramus Educational [5]. Программа является бесплатной и урезанной версией программы Ramus. Не смотря на то что Ramus Educational не поддерживает навигацию по модели и формирование отчетов, программа подходит для создания небольших диаграмм. Ramus Educational предоставляет следующие возможности:

* создание диаграмм IDEF0;
* создание диаграмм DFD;
* создание классификаторов (систематизированный, иерархический перечень наименований объектов с присвоенными кодами, описание объекта с помощью атрибутов);
* экспорт/импорт в файлы формата IDL. таким образом реализуя частичную совместимость с подобными программами (например, с CA Erwin Process Modeler, в том числе Ramus полной версии).

При разработке системы сжатия будет использоваться технология Java. Программы созданные на Java компилируются в байт код и выполняются на виртуальной машине, что дает им кросс-платформенность на уровне бинарных сборок. Java предоставляет огромное количество стандартных библиотек, в том числе и для создания графического пользовательского интерфейса, работой с файловой системой, и тому подобное. Написание программного кода будет выполняться в интегрированной среде разработки программного обеспечения IntelliJ IDEA. Программа предоставляется различные плагины для обеспечения более удобного оперирования с программным кодом. Также в IntelliJ IDEA включены средства для интеграции с [системами управления версиями](https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0_%D1%83%D0%BF%D1%80%D0%B0%D0%B2%D0%BB%D0%B5%D0%BD%D0%B8%D1%8F_%D0%B2%D0%B5%D1%80%D1%81%D0%B8%D1%8F%D0%BC%D0%B8) и разнообразные инструменты для упрощения конструирования [графического интерфейса пользователя](https://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D1%84%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9_%D0%B8%D0%BD%D1%82%D0%B5%D1%80%D1%84%D0%B5%D0%B9%D1%81_%D0%BF%D0%BE%D0%BB%D1%8C%D0%B7%D0%BE%D0%B2%D0%B0%D1%82%D0%B5%D0%BB%D1%8F).

На рисунке 2.3 представлены методологии, технологии и средства реализации, которые будут использоваться на каждом этапе жизненного цикла при проектировании ССБД.

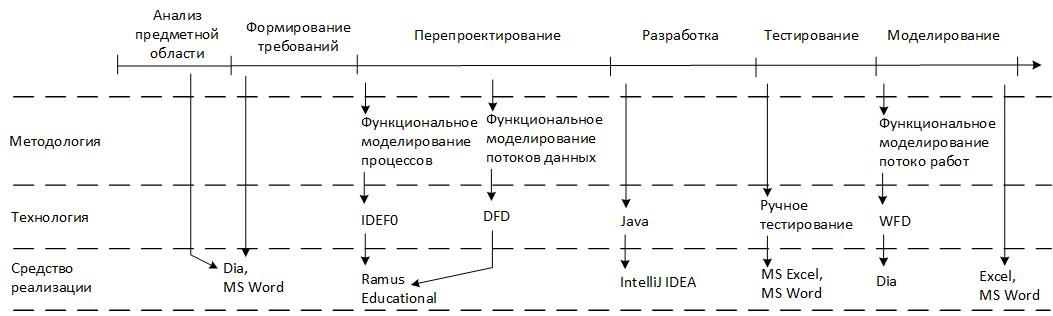


Рис. 2.3. Комплекс технологий используемый при проектировании ССБД.

# **2.4 Формализованное описание комплекса технологий при проектировании ИС**

На основе схемы приведенной на рисунке 2.3 можно формализовать описание комплекса технологий. Формализованное описание комплекта технологий при проектировании ИС представлено формулой 2.1:





,(2.1)

где  – технологии проектирования;

 – средства анализа предметной области;

 – средства анализа требований;

 – средства анализа программного обеспечения;

 – средства описания предметной области;

 – средства описания магасистемы;

 – средства проектирования декомпозиции процесса;

 – средства проектирования потока данных;

 – средства проектирования потоков работ;

 – средства разработки программного обеспечения;

 – средства тестирования программного обеспечения;

 – средства моделирования результатов эксперимента.

При анализе предметной области, требований и программного обеспечения использовался MS Word. Анализ требований основывался исходя из анализа созданных программ сжатия данных и анализа предметной области. Описания предметной области и мегасистемы выполнены в виде схемы и реализованы в Dia. Пояснения к схемам выполнены в MS Word. При выполнении проектирования декомпозиции процесса использовалась технология функционального моделирования процессов (IDEF0). Описание процессов выполнялось на основе методологии функционального моделирования потоков данных (DFD) и потоков работ (WFD). Для построения диаграмм IDEF0 и DFD используется CASE-средство Ramus Educational. Создание схем, которые позволяю наглядно представить структуру программы, будут выполнены в Dia. Разработка программного обеспечения будет выполнятся в интегрированной среде разработки IntelliJ IDEA. Тестирование разработанного программного продукта будет выполнено вручную, результаты тестирования будут записаны в MS Excel. При моделировании результатов эксперимента будет использоваться MS Excel, в котором будут отображены результаты экспериментов в табличном виде и на их основе построены графики.

# **3 Оценка эффективности выбранного комплекса технологий при проектировании ИС**

# **3.1 Выбор критериев оценки комплекса технологий при проектировании ИС**

Критерии, которые используются для оценки выбранного комплекса технологий, представлены в таблице 3.1.

Таблица 3.1. Критериев оценки комплекса технологий

|  |  |  |
| --- | --- | --- |
| № | Критерии | Комментарий |
| 1 | Группа критериев доступности технологий по лицензированию, Q1 |  |
| 1.1 | Лицензирование, q1 | Все использованное ПО лицензированное |
| 1.2 | В свободном доступе, q2 | У каждого программного обеспечения, которое используется в работе, можно найти бесплатную версию (возможно с урезанным функционалом, но достаточным для реализации поставленной задачи) |
| 2 | Группа критериев применимости, Q2 |  |
| 2.1 | Применимо к нескольким этапам, q3 | Все ПО, кроме Excel, используются на нескольких  этапах, например, Ramus Educational, Dia |
| 2.2 | Автоматизированость этапов проектирования, q4 | Ramus Educational обладает автоматизированностью при создании диаграммы IDEF0 (декомпозиция) |

Возможно применение других критериев, например, доступность реализации, сложности ПО, но так как предложенные технологии проектирования ориентированы на специалистов в области ИС и не требуют дополнительных навыков для работы с ними, поэтому доступность по применимости не рассматривается.

Комплекс технологий, который будет использоваться в научно-исследовательской работе включает в себя следующие программные обеспечения: MS Word, MS Excel, Dia, Ramus Educational, IDE IntelliJ IDEA.

# **3.2 Расчет показателя эффективности оценки комплекса технологий**

Приведем критерии описанные в пункте 3.1 к одному направлению экстремизации.

Группа критериев доступности технологий, Q1:

* вероятность того, что ПО лицензированное, q1 (критерий максимизируется);
* вероятность того, что ПО распространяется в свободном доступе, q2 (критерий максимизируется).

Группа критериев применимости, Q2:

* вероятность того, что используемое ПО применимо к нескольким этапам, q3 (критерий максимизируется);
* вероятность того, что используемое ПО поддерживает автоматизированость этапов проектирования, q4 (критерий максимизируется).

Сформированная иерархия изображена на рисунке 3.1.

Расчет показателя эффективности оценки комплекса технологий выполняется методом иерархической свертки [1]. В Excel выполнен расчет эффективности выбранного комплекса технологий. В таблице 3.2 приведены значения критериев.

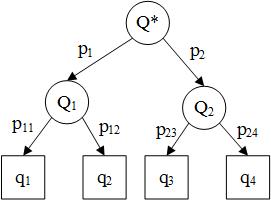


Рис. 3.1. Иерархия критериев оценки комплекса технологий.

Таблица 3.2. Значения критериев оценки комплекса технологий

|  |  |
| --- | --- |
| Обозначение | Значение |
| q1 | 0,8 |
| q2 | 0,9 |
| q3 | 0,8 |
| q4 | 0,5 |

В таблице 3.3 приведены значения коэффициентов приоритета.

Таблица 3.3. Значения коэффициентов приоритета

|  |  |
| --- | --- |
| Обозначение | Значение |
| p11 | 0,4 |
| p12 | 0,6 |
| p23 | 0,6 |
| p24 | 0,4 |
| p1 | 0,4 |
| P2 | 0,6 |

В таблице 3.4 отображены результаты расчетов.

Таблица 3.4. Результаты расчетов

|  |  |
| --- | --- |
| Обозначение | Значение |
| Q1 | 0,875 |
| Q2 | 0,737 |
| Q\* | 0,81 |

По результатам расчетов Q\* равна 0,81, значит выбранный комплекс технологий обладает высоким показателем эффективности.

Для сравнения выполним расчет показателя эффективности для аналогов программного обеспечения входящего в выбранный комплекс технологий. Для расчетов примем, что в комплекс технологий вместо Dia входит платный аналог Visio, вместо Ramus Educational – аналог Ramus. В этом случае значения критериев группы Q1 будет ниже. В таблице 3.5 приведены значения критериев для аналогов выбранному комплексу технологий.

Таблица 3.5. Значения критериев для аналогов выбранному комплексу технологий

|  |  |
| --- | --- |
| Обозначение | Значение |
| q1 | 0,6 |
| q2 | 0,5 |
| q3 | 0,8 |
| q4 | 0,5 |

В таблице 3.6 отображены результаты расчетов для аналогов выбранному комплексу технологий.

Таблица 3.6. Результаты расчетов для аналогов выбранному комплексу технологий

|  |  |
| --- | --- |
| Обозначение | Значение |
| Q1 | 0,545 |
| Q2 | 0,737 |
| Q\* | 0,684 |

По результатам расчетов Q\* равна 0,684, значит при использовании аналогов выбранным программным обеспечениям, полученный комплекс технологий будет обладает хорошим показателем эффективности. Однако, у выбранного комплекса технологий показатель эффективности выше.

# **4 Оценка возможных эксплуатационных характеристик ИС задаче НИР**

# **4.1 Расчет информационных характеристик системы (требуемый объем жесткого диска при учете интенсивности приращения информационного фонда)**

Рассчитаем длину логической записи файла, который составляется при анализе данных, используя следующую формулу:

, (4.1)

где Mj – число групп полей в записях;

lij – длина группы.



Объем памяти, необходимый для размещения информационного фонда без учета системных данных и указателей рассчитывается по формуле:

 , (4.2)

где R – число отношений в РБД;

lij – количество записей j-го файла.



Количество обращений к логическим записям определяется по формуле:

 , (4.3)

где – количество обращений к записям j-го типа в i-м запросе.

Запись j-го типа одна, запрос может быть либо на чтение, либо на запись.

Интенсивность обращения к информационному фонду рассчитывается по формуле:

 , (4.4)

где  – частота выполнения i-го запроса

– число запросов, обработка которых предусмотрена СУБД.

Частота выполнения операции записи равна 0,2; частота выполнения операции чтения 0,8, значит интенсивность обращения в информационному фонду равна:

.

Средний объем данных, предоставляемый пользователю во время выполнения i-ого запроса определяется по следующей формуле:

. (5)

На основании ранее выполненных расчетов, получим средний объем данных, предоставляемый пользователю во время выполнения i-ого запроса равный:

.

# **БИБЛИОГРАФИЧЕСКИЙ СПИСОК**

1. Воронин А.Н. Вложенные скалярные свертки векторного критерия / А. Н. Воронин // Проблемы управления и информатики. – 2003. – № 5. – С. 10 – 21.

2. Игнатьев А.В. Методы и средства проектирования ИС и ИТ: эл. учеб. / А.В. Игнатьев, 2014. – 56 с.

3. Орлов, С.А. Технологии разработки программного обеспечения: учеб. / С.А. Орлов. – СПб.: Питер, 2002. – 464 с.

4. Шпилина Д.Ю. Автоматизация банковской деятельности // Экономика и менеджмент инновационных технологий. 2014. № 1 [Электронный ресурс]. URL: http://ekonomika.snauka.ru/2014/01/3591 (дата обращения: 20.11.2016).

5. Ramus [Электронный ресурс]. URL: <http://ramussoftware.com> (дата обращения: 18.02.2017).