

## Содержание

Введение . . . . .	4
Краткое описание предметной области . . . . .	4
Существующие аналоги . . . . .	4
Описание системы . . . . .	5
Система «А» . . . . .	6
Система «Б» . . . . .	6
Система «В» . . . . .	6
Основания для разработки . . . . .	7
Назначение разработки . . . . .	7
1 Аналитический раздел . . . . .	8
1.1 Структура почтовой системы . . . . .	8
1.2 Типы МТА . . . . .	10
1.3 Стандарты, регламентирующие передачу сообщений по сети Интернет . . . . .	11
1.4 Существующие подходы к созданию всячины . . . . .	11
2 Конструкторский раздел . . . . .	14
2.1 Архитектура всячины . . . . .	14
2.2 Подсистема всякой ерунды . . . . .	14
2.2.1 Блок-схема всякой ерунды . . . . .	14
3 Технологический раздел . . . . .	16
4 Экспериментальный раздел . . . . .	18
Заключение . . . . .	19
Список использованных источников . . . . .	20
А Картинки . . . . .	21
Б Еще картинки . . . . .	22

## Глоссарий

**Спам** — Анонимные не запрошенные массовые рассылки электронной почты, то есть электронный эквивалент бумажной рекламной корреспонденции, засоряющей обычные почтовые ящики.

## Обозначения и сокращения

— SMTP — англ. Simple Mail Transfer Protocol — простой протокол передачи почты. Сетевой протокол, предназначенный для передачи электронной почты в сетях TCP/IP

— POP3 — англ. Post Office Protocol Version 3 — протокол почтового отделения, версия 3. Используется почтовым клиентом для получения сообщений электронной почты с сервера.

— IMAP — англ. Internet Message Access Protocol — протокол прикладного уровня для доступа к электронной почте.

## **Введение**

Настоящее техническое задание разработано в рамках учебной программы по курсу «Технология программирования» на программное изделие «Распределенная система обнаружения и фильтрации спама в протоколах POP3, IMAP, SMTP».

### **Краткое описание предметной области**

За последние десять лет сфера применения спама расширилась, а объем доставки - вырос значительно. Первое время спам рассылался напрямую на единичные адреса пользователей, и его было легко блокировать. Со временем появились высокоскоростные интернет-каналы, которые дали быструю и дешевую возможность массово рассылать спам-сообщения. Модемы пользователей не оснащались средствами защиты от несанкционированного доступа и могли использоваться злоумышленниками из любой точки планеты. Другими словами, модемы ничего не подозревающих пользователей рассылали огромные объемы спама.

Так продолжалось, пока производители аппаратного обеспечения не научились оснащать оборудование средствами защиты от спама, а спам-фильтры не стали более эффективными. Однако спам тоже эволюционировал: усовершенствовались не только способы рассылки, но и приемы, помогающие злоумышленникам обойти спам-фильтры.

В ходе анализа предметной области были рассмотрены наиболее популярные спам-фильтры и другие механизмы, применяющиеся для фильтрации потока электронной почты от спама.

### **Существующие аналоги**

В рамках настоящей работы был произведен анализ рынка программных продуктов, позволяющих фильтровать электронную почту от спама. Было выявлено, что существует множество решений позволяющих осуществлять определение нежелательной почты для конкретного пользователя.

Существует два класса подобных продуктов:

- а) спам-фильтры на почтовых клиентах;
- б) спам-фильтры на почтовых серверах.

Спам-фильтры почтовых клиентов основаны на обработке входящих сообщений согласно заданным пользовательским правилам. Почтовые сервера используют так называемые черные и серые списки доверенных адресов входящей почты.

Таким образом, распределенных систем, позволяющих осуществлять фильтрацию спама обнаружено не было.

## Описание системы

Назначением разрабатываемой системы является осуществление фильтрации спама, которая основывается на информации от многих пользователей, которые имеют почтовые ящики на различных почтовых серверах.

На рисунке 0.1 представлена структура предметной области и ее составные части:

- а) система «А», почтовый клиент (MUA);
- б) система «Б», почтовый сервер (MTA);
- в) система «С», анти-спам сервер.

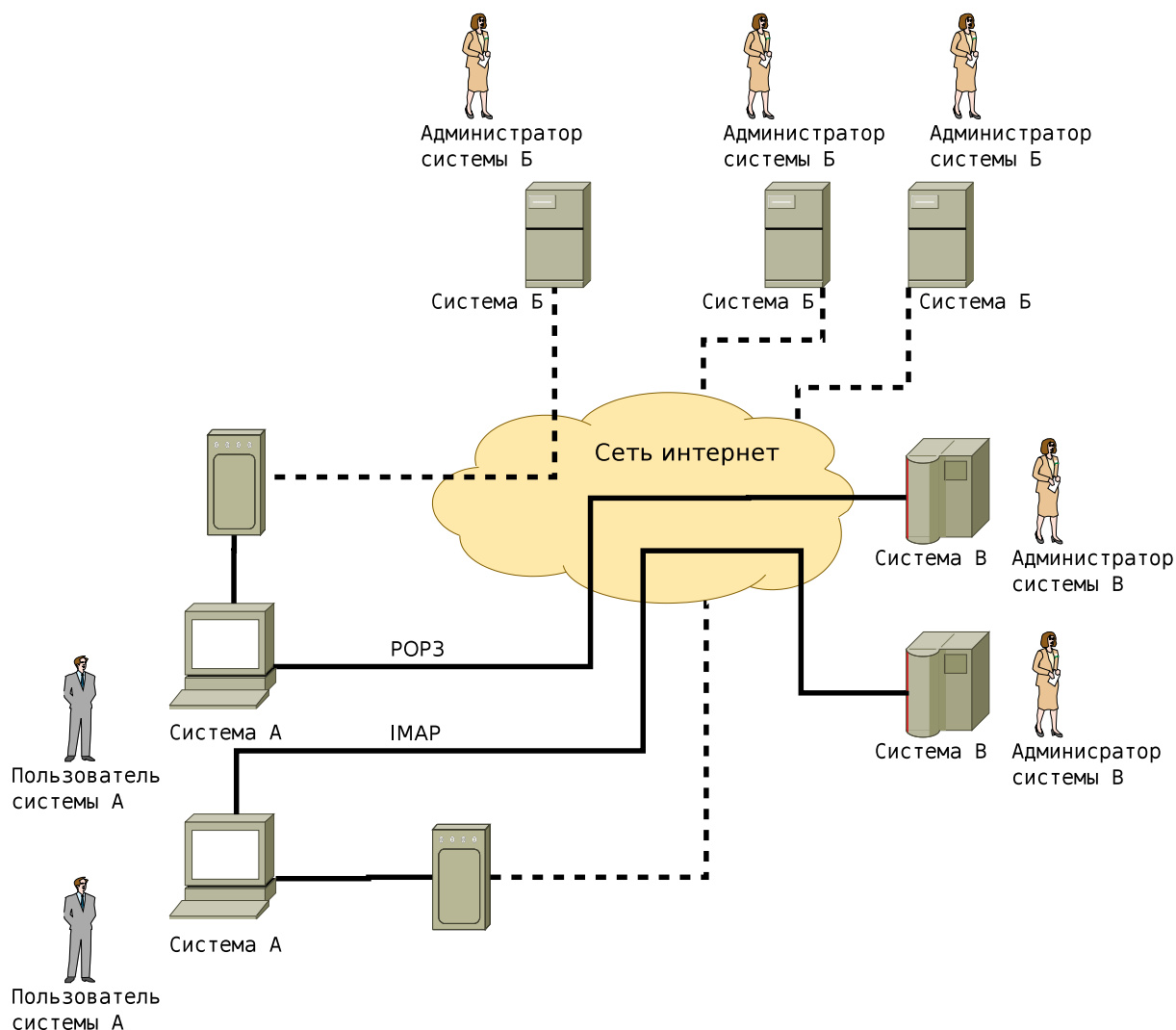


Рисунок 0.1 — Структура предметной области

Перечисленные системы в дальнейшем будут также называться субъектами РСОИ или системами-участниками.

### **Система «А»**

Целью систем данного типа является предоставление удобного пользовательского интерфейса для работы с электронной почтой. С точки зрения почтовой системы представляют собой MUA — программное обеспечение, устанавливаемое на компьютере пользователя и предназначенное для получения, написания, отправки и хранения сообщений электронной почты одного или нескольких пользователей (в случае, например, нескольких учётных записей на одном компьютере) или нескольких учётных записей одного пользователя.

Системы «А» бывают двух типов в зависимости от протокола получения входящих сообщений:

- а) POP3;
- б) IMAP.

Протокол POP3 подразумевает передачу входящих сообщений почтовым сервером и сохранение электронных писем на локальном компьютере пользователя. При использовании POP3 клиент подключается к серверу только на промежуток времени, необходимый для загрузки новых сообщений. При использовании IMAP соединение не разрывается, пока пользовательский интерфейс активен, а сообщения загружаются только по требованию клиента. Это позволяет уменьшить время отклика для пользователей, в чьих ящиках имеется много сообщений большого объёма. Протокол POP требует, чтоб текущий клиент был единственным подключенным к ящику. IMAP позволяет одновременный доступ нескольких клиентов к ящику и предоставляет клиенту возможность отслеживать изменения, вносимые другими клиентами, подключенными одновременно с ним.

### **Система «Б»**

Системы аккумулируют информацию необходимую для фильтрации электронной почты пользователя от спама.

### **Система «В»**

Системы данного типа отвечает за отправку почты и представляют собой почтовые сервера, которые обычно выполняют роль MTA и MDA. Некоторые почтовые сервера (программы) выполняют роль как MTA, так и MDA, некоторые подразумевают разделение на два независимых сервера: сервер-MTA и сервер-MDA (при этом, если для доступа к почтовому ящику используются различные протоколы — например, POP3 и IMAP, — то MDA в свою очередь может быть реализован либо как единое приложение, либо как набор приложений, каждое из которых отвечает за отдельный протокол).

## **Основания для разработки**

Разработка ведется в рамках выполнения лабораторных работ по курсу «Технология программирования» и курсового проекта по курсу «Распределенные системы обработки информации» на основании учебного плана МГТУ им. Баумана на 12-й семестр для факультета ИУ.

## **Назначение разработки**

Представленные системы, позволяющие фильтровать почту от нежелательных сообщений, основываются на спам-фильтрах, которые имеют следующие недостатки:

- а) необходимо обучение;
- б) история обучения локальная и относится к конкретному пользователю.

Настоящая разработка должна обеспечить передачу информации между пользователями о признаках обнаруженного спама. Распределенная система обнаружения спама позволит сократить время на обучение спам-фильтров для отдельных пользователей, тем самым повысит удобство использования почтовых клиентов.

## 1 Аналитический раздел

В данном разделе анализируется и классифицируется существующая всячина и пути создания новой всячины. А вот отступ справа в 1 см. — это хоть и по ГОСТ, но ведь диагноз же...

### 1.1 Структура почтовой системы

Почтовая система в общем случае может быть представлена как совокупность следующих элементов:

- MUA — (англ. Mail User Agent) почтовый агент пользователя, почтовый клиент);
- MTA — (англ. Mail Transfer Agent) агент пересылки почты;
- MDA — (англ. Mail Delivery Agent) агент доставки почты.

MUA представляет собой программное обеспечение, устанавливаемое на компьютере пользователя и предназначенное для получения, написания, отправки и хранения сообщений электронной почты одного или нескольких пользователей (в случае, например, нескольких учётных записей на одном компьютере) или нескольких учётных записей одного пользователя. Предназначением является предоставление удобного пользовательского интерфейса для работы с электронной почтой.

MTA отвечает за отправку почты. MDA отвечает за доставку почты конечному пользователю. Почтовые сервера обычно выполняют роль MTA и MDA. Некоторые почтовые сервера (программы) выполняют роль как MTA, так и MDA, некоторые подразумевают разделение на два независимых сервера: сервер-MTA и сервер-MDA (при этом, если для доступа к почтовому ящику используются различные протоколы — например, POP3 и IMAP, — то MDA в свою очередь может быть реализован либо как единое приложение, либо как набор приложений, каждое из которых отвечает за отдельный протокол).

Только один MTA может функционировать на одной рабочей станции, так как исключительно одно приложение может быть назначено для получения входящих сообщений от других рабочих станций. Как правило, это привелигированная программа, которая прослушивает входящие TCP/IP соединения по SMTP-порту и имеет возможность сохранять данные в пользовательские почтовые ящики.

MTA способен принимать множество сообщений одновременно. Если в связи с непредвиденными обстоятельствами MTA не может доставить сообщение конечному пользователю, то посылается сообщение с причиной неудачной отправки. MTA хранит все сообщения, которые не могут быть своевременно отправлены конечному пользователю. Через определенные промежутки времени MTA инициирует повторную



отправку подобных сообщений. Чаще всего невозможность доставки почты связана с проблемами сетевого соединения и с отключением целевой рабочей станции.

С точки зрения МТА существуют два типа источников входящих сообщений: локальные процессы и другие рабочие станции. Также можно выделить три типа адресатов: локальные файлы, локальные процессы и другие рабочие станции.

Выделение агентов МТА и MUA означает, что они могут функционировать на различных рабочих станциях.

В верхней части рисунка ?? MUA, МТА и дисковое хранилище являются частью единой системы, которая выделена штриховой линией. Пользователи получают доступ к системе путем аутентификации и авторизации с помощью ввода логина и соответствующего пароля. MUA запускается с помощью пользовательской команды, как процесс операционной системы, и когда инициирует передачу сообщения МТА для последующей отправки, начинает взаимодействие с другим процессом операционной системы. MUA и МТА взаимодействуют с авторизованным пользователем, поэтому МТА обычно производит проверку того, что идентификационная информация пользователя включена в исходящее сообщение. Как описано в RFC 822, если в теле сообщения не указывается адресант From, то МТА обязан добавить запись Sender с указанием идентификационной информации отправителя.

МТА хранит сообщения в так называемой области спулинга. Под областью спулинга будем понимать дисковое хранилище, которое использует МТА для временного хранения очереди сообщений перед их отправкой.

Сообщения, которые предназначены другим удаленным рабочим станциям передаются по сети Интернет другим МТА с использованием протокола Simple Mail Transfer Protocol (SMTP). В том случае когда сервер-отправитель и сервер-получатель оба напрямую подключены к сети Интернет, сообщение может быть доставлено напрямую от отправителя к получателю. Однако иногда сообщению приходится преодолевать маршруты через промежуточные МТА. Большие организации часто организуют свои почтовые системы таким образом, что все входящие сообщения поступают на главный почтовый сервер. После этого сообщения пересылаются через другие серверы локальной сети. Когда сообщение доставляется получателю, МТА сохраняет его в почтовом ящике пользователя, который затем с помощью MUA получает доступ к своей почте.

Также промежуточные МТА используются в тех случаях, когда целевой почтовый сервер не доступен или сетевое соединение не может быть установлено. Преимущество такого подхода заключается в том, что сообщения аккумулируются на серверах максимально приближенных к целевому адресату и могут быть своевременно доставлены.

Нижняя часть рисунка иллюстрирует такую конфигурацию почтовой подсистемы получателя, когда MUA и МТА располагаются на различных рабочих станци-

ях. Такая конфигурация позволяет разделить процессы получения и отправки писем на отдельные операции. В процессе чтения почты пользователь напрямую взаимодействует с MUA, который использует протоколы POP3 (RFC 1939) или IMAP (RFC 2060) для получения доступа к почтовым ящикам пользователя и удаленным папкам на сервере. Для того чтобы осуществить подобные операции, пользователь должен быть авторизован в почтовой системе. Однако, протоколы POP3 и IMAP не содержат средств для пересылки сообщений. MUA подобного типа используют протокол SMTP для отправки сообщения MTA. Таким образом протокол SMTP, который изначально описывал передачу сообщений между MTA, в настоящее время используется также для транзита сообщений от MUA к MTA. Такое использование приводит к ряду проблем:

- MTA не может различить сообщения, поступающие от других MTA, от сообщений, предоставляемых MUA.
- Пользователь, отправляющий почту, не является установленным, то есть прошедшим авторизацию. В связи с этим MTA не предоставляется возможным определить является ли домен отправителя реально существующим.
- MUA может использовать различные сервера для отправки почты. Также специализированные MUA могут отправлять сообщения адресату напрямую через сеть интернет. Поэтому возможна ситуация, когда злоумышленники пытаются отправить поток нежелательной почты на произвольные сервера для ретрансляции.

На данный момент разработаны протоколы, которые способны решать описанные проблемы. Однако широкого применения они не получили.

## **1.2 Типы MTA**

В простейшем случае отдельные рабочие станции или сервера в небольших офисах и домах, которые работают с несколькими почтовыми ящиками в одном домене, получают входящие сообщения от одного провайдера интернет услуг и передают сообщения провайдеру для последующей доставки адресату. Среди хостов, имеющих постоянное подключение к сети интернет, также имеются хосты, которые периодически подключаются к интернету для получения входящей почты с сервера и отправки исходящей почты.

Рабочие станции, которые имеют постоянное подключение к сети интернет, могут не посылать сообщения через один и тот же сервер.

### 1.3 Стандарты, регламентирующие передачу сообщений по сети Интернет

Передача сообщений по сети интернет регламентируется согласно стандарту RFC 822, в котором определяется формат передаваемых сообщений. Протокол SMTP обмена сообщениями между хостами описывается в стандартах RFC 821, RFC 1123, а также в некоторых других стандартах, определяющих расширения протокола SMTP.

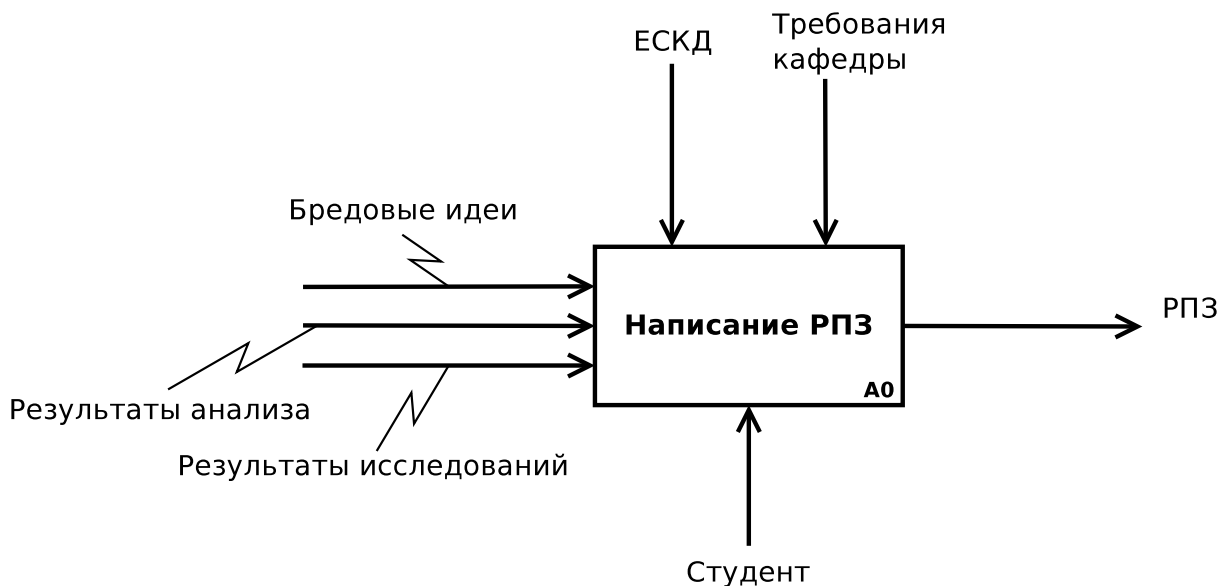


Рисунок 1.1 — Рисунок

В [1] указано, что...

Кстати, про картинки. Во-первых, для фигур следует использовать `[ht]`. Если и после этого картинки вставляются «не по ГОСТ», т.е. слишком далеко от места ссылки, — значит у вас в РПЗ **слишком мало текста**! Хотя и ужасный параметр `!ht` у окружения `figure` тоже никто не отменял, только при его использовании документ получается страшный, как в ворде, поэтому просьба так не делать по возможности.

## 1.4 Существующие подходы к созданию всячины

Известны следующие подходы...

- а) Перечисление с номерами.
- б) Номера первого уровня. Да, ГОСТ требует именно так — сначала буквы, на втором уровне — цифры. Чуть ниже будет вариант «нормальной» нумерации и советы по её изменению. Да, мне так нравится: на первом уровне выравнивание элементов как у обычных абзацев. Проверим теперь вложенные списки.
  - 1) Номера второго уровня.
  - 2) Номера второго уровня. Проверяем на длииииной-преддддддддддддддддддддд строке, что получается.... Сойдёт.

в) По мнению Лукьяненко, человеческий мозг старается подвести любую проблему к выбору из трех вариантов.

г) Четвёртый (и последний) элемент списка.

Теперь мы покажем, как изменить нумерацию на «нормальную», если вам этого захочется. Пара команд в начале документа поможет нам.

1) Изменим нумерацию на более привычную...

2) ... нарушим этим гост.

а) Но, пожалуй, так лучше.

В заключение покажем произвольные маркеры в списках. Для них нужен пакет **enumerate**.

1. Маркер с арабской цифрой и с точкой.

2. Маркер с арабской цифрой и с точкой.

I. Римская цифра с точкой.

II. Римская цифра с точкой.

В отчётах могут быть и таблицы — см. табл. 1.1 и 1.2. Небольшая таблица делается при помощи **tabular** внутри **table** (последний полностью аналогичен **figure**, но добавляет другую подпись).

Таблица 1.1 — Пример короткой таблицы с длинным названием на много длинных-длинных строк

Тело	$F$	$V$	$E$	$F + V - E - 2$
Тетраэдр	4	4	6	0
Куб	6	8	12	0
Октаэдр	8	6	12	0
Додекаэдр	20	12	30	0
Икосаэдр	12	20	30	0
Эйлер	666	9000	42	$+\infty$

Для больших таблиц следует использовать пакет **longtable**, позволяющий создавать таблицы на несколько страниц по ГОСТ.

Для того, чтобы длинный текст разбивался на много строк в пределах одной ячейки, надо в качестве ее формата задавать **p** и указывать явно ширину: в мм/дюймах (**110mm**), относительно ширины страницы (**0.22\textwidth**) и т.п.

Можно также использовать уменьшенный шрифт — но, пожалуйста, тогда уж во **всей** таблице сразу.

Таблица 1.2 — Пример длинной таблицы с длинным названием на много длинных-длинных строк

Вид шума	Громкость, дБ	Комментарий
Порог слышимости	0	
Шепот в тихой библиотеке	30	Конечно, это было до эпохи мобильных (внутри машины)
Обычный разговор	60-70	
Звонок телефона	80	
Уличный шум	85	
Гудок поезда	90	
Шум электрички	95	
Порог здоровой нормы	90-95	Длительное пребывание на более громком шуме может привести к ухудшению слуха
Мотоцикл	100	(модель бензокосилки) (Doom в целом вреден для здоровья)
Power Mower	107	
Бензопила	110	
Рок-концерт	115	
Порог боли	125	feel the pain
Клепальный молоток	125	(автор сам не знает, что это)
Порог опасности	140	Даже кратковременное пребывание на шуме большего уровня может привести к необратимым последствиям
Реактивный двигатель	140	Необратимое полное повреждение слуховых органов Интересно, почему?..
	180	
Самый громкий возможный звук	194	

## 2 Конструкторский раздел

В данном разделе проектируется новая всячина.

### 2.1 Архитектура всячины

**Проверка** параграфа. Вроде работает.

**Вторая проверка** параграфа. Опять работает.

Вот.

- Это список с «палочками».
- Хотя он и не по ГОСТ, кажется.

1) Поэтому для списка, начинающегося с заглавной буквы, лучше список с цифрами.

Формула 2.1 совершенно бессмысленна.

$$a = cb \tag{2.1}$$

Окружение `cases` опять работает (см. 2.2), спасибо И. Короткову за исправления..

$$a = \begin{cases} 3x + 5y + z, & \text{если хорошо} \\ 7x - 2y + 4z, & \text{если плохо} \\ -6x + 3y + 2z, & \text{если совсем плохо} \end{cases} \tag{2.2}$$

### 2.2 Подсистема всякой ерунды

Культурная вставка dot-файлов через утилиту `dot2tex` (рис. 2.1).

#### 2.2.1 Блок-схема всякой ерунды

**Кстати о заголовках**

У нас есть и **subsubsection**. Только лучше её не нумеровать.

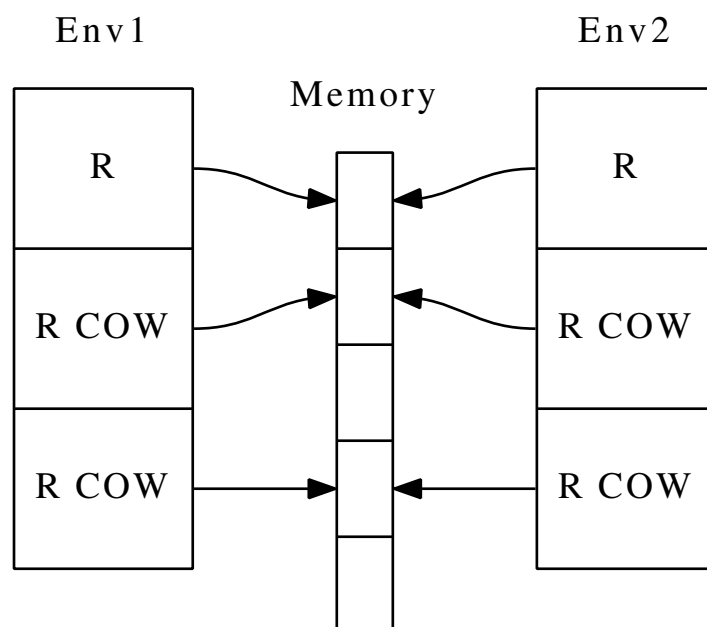


Рисунок 2.1 — Рисунок

### 3 Технологический раздел

В данном разделе описано изготовление и требование всячины. Благодаря пакет **underscore** эскейпить подчёркивание не нужно (**some\_function**).

Для вставки кода есть пакет **listings**. К сожалению, пакет **listings** всё ещё работает криво при появлении в листинге русских букв и кодировке исходников utf-8. В данном примере он (увы) на лету конвертируется в koi-8 в ходе сборки pdf.

Есть альтернатива **listingsutf8**, однако она работает лишь с `\lstinputlisting`, но не с окружением **lstlisting**

Вот так можно вставлять псевдокод (питоноподобный язык определен в шаблоне):

Листинг 3.1 — Алгоритм оценки дипломных работ

```
1 def EvaluateDiplomas():
2     for each student in Masters:
3         student.Mark ← 5
4     for each student in Engineers:
5         if Good(student):
6             student.Mark ← 5
7         else:
8             student.Mark ← 4
```

Еще в шаблоне определен псевдоязык для BNF:

Листинг 3.2 — Грамматика

```
1 ifstmt → "if" "(" expression ")" stmt |
2         "if" "(" expression ")" stmt1 "else" stmt2
3 number → digit digit *
```

В листинге 3.3 работают русские буквы. Сильная магия. Однако, работает только во включаемых файлах, прямо в `TeX` нельзя.

Листинг 3.3 — Пример (**test.c**)

```
1 #include <stdio.h>
2
3 int main()
4 {
5     /* Комментарий на русском с пробелами */
6     printf("Это строка с пробелами и русскими буквами");
7
8     return 0;
9 }
```



Можно также использовать окружение **verbatim**, если **listings** чем-то не устраивает. Только следует помнить, что табы в нём «съедаются». Существует также команда **verbatiminput** для вставки файла.

```
a_b = a + b; // русский комментарий
if (a_b > 0)
    a_b = 0;
```

## 4 Экспериментальный раздел

В данном разделе проводятся вычислительные эксперименты. А на рис. 4.1 показана схема мыслительного процесса автора...

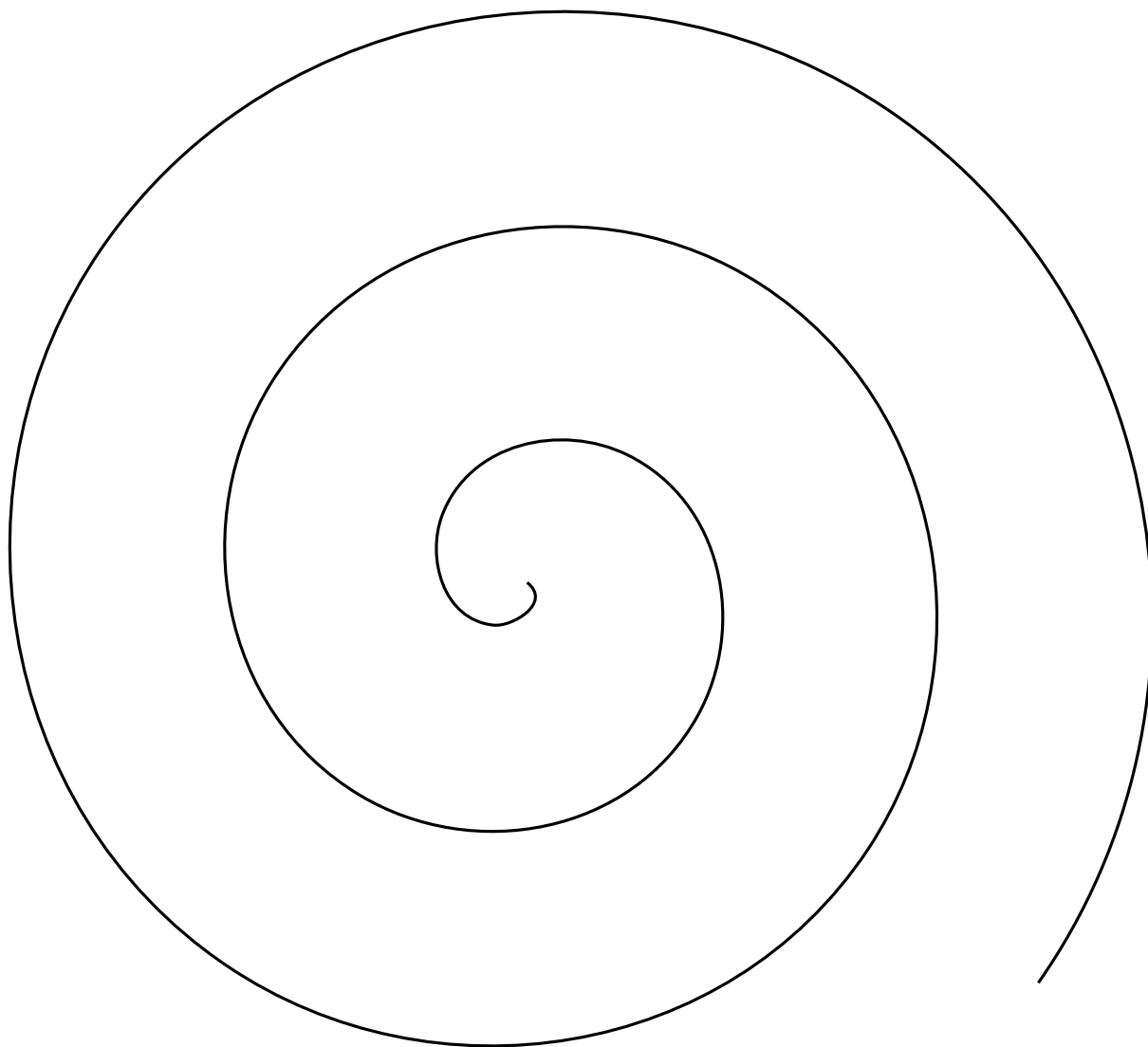


Рисунок 4.1 — Как страшно жить

## Заключение

В результате проделанной работы стало ясно, что ничего не ясно...

## Список использованных источников

1. *Пупкин, Василий*. L<sup>A</sup>T<sub>E</sub>X для «чайников» / Василий Пупкин, А. Эйнштейн.  
— М., 2009.

## Приложение А    Картинки

Рисунок А.1 — Картинка в приложении. Страшная и ужасная.

## Приложение Б    Еще картинки

Рисунок Б.1 — Еще одна картинка, ничем не лучше предыдущей. Но надо же как-то заполнить место.