

УТВЕРЖДАЮ

Директор ООО «Центр-инвест ИТ»

_____ Дорошкевич Е. В.

«__» _____ 2025 г.

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ «СМС-Анализатор»

Руководство администратора/пользователя

Описание функциональных характеристик и информация, необходимая для
установки и эксплуатации ПО «СМС-Анализатор»

Оглавление

1.	Термины.....	3
2.	Описание ПО «СМС-Анализатор».....	3
3.	Программные и аппаратные требования к системе.....	4
4.	Руководство администратора/пользователя	5
4.1.	Установка	5
4.2.	Основные компоненты.....	5
4.3.	Результаты обработки шаблонов	6
5.	Алгоритм составления дерева поиска	6
6.	Ограничения.....	10
7.	Авторские права	10

1. Термины

ПО «СМС-Анализатор» — это программное обеспечение, предназначенное для анализа текстовых сообщений, сопоставления их с подходящим шаблоном и последующей записи результата в базе данных (далее – БД).

Сообщение — текстовый файл, содержащий информацию, подлежащую анализу.

Шаблон — образец текста, с которым сравниваются сообщения, состоящий из определенных элементов, которые несут под собой возможный набор слов или символов.

Компьютер-клиент (Устройство)— компьютер, в качестве которого может выступать рабочая станция, ноутбук или любое другое цифровое устройство, включающее системное программное обеспечение такого компьютера/устройства (операционную систему, драйвера для работы с техническим обеспечением, веб-браузер и т.п.), подключенное к локальной сети и (или) сети связи общего пользования, на который загружается и на котором запускается и работает клиентская часть ПО.

2. Описание ПО «СМС-Анализатор»

ПО «СМС-Анализатор» представляет собой комплекс, состоящий из серверной части, запускаемой на компьютере и базы данных для хранения сообщений, шаблонов и результатов анализа.

Сервис запрашивает из базы данных пакет СМС и список шаблонов, из полученных шаблонов строит деревья поиска, сопоставляет с ними каждое СМС и сохраняет результат обратно в базу данных.

Примеры использования:

- Использование для анализа обращений клиентов и определения их потребностей
- Использование для сбора статистики с целью оптимизации расходов на отправку сообщений
- Сопоставление результатов анализа с данными операторов

ПО «СМС-Анализатор» состоит из следующих элементов:

- sms-analyzer - запрашивает данные из таблиц с шаблонами и сообщениями в базе данных, анализирует их и отправляет результат в таблицу результатов;
- база данных Postgres - хранит шаблоны, СМС и таблицу соответствий (id сообщения → id шаблона).

В разработке ПО «СМС-Анализатор» использовались:

- технологии и средства разработки: IntelliJ Idea, JDK, СУБД Postgres;
- языки программирования: Kotlin, SQL.

3. Программные и аппаратные требования к системе

Для корректной работы с ПО «СМС-Анализатор» необходима следующая конфигурация рабочего места администратора.

№ n/n	Параметр	Роль	Минимально допустимые значения	Рекомендуемые значения
1	2	3	4	5
Общие требования				
1.	Компьютер-клиент: • Процессор; • Оперативная память • Видеоадаптер • Свободное место на жестком диске	Устройство взаимодействия Пользователя с ПО	• Intel Core i3 6-го поколения или аналогичный • 4 ГБ • Intel HD Graphics 530 и выше • 64 ГБ	Intel Core i5 6-го поколения или аналогичный 8 ГБ
2.	Системное ПО компьютера-клиента: • Операционная система; • Консоль или интерфейс СУБД	Обеспечивает возможность запуска и среду для работы пользователя ПО	Взаимодействие с клиентом осуществляется через интерфейс СУБД. Кроссплатформенность используемых технологий позволяет клиентской части ПО работать в различных операционных системах. Например, • Red Hat Enterprise Linux (RHEL) (32-bit or 64-bit); • Novel SUSE Linux Enterprise Server (SLES) (32-bit or 64-bit); • Microsoft Windows (32-bit or 64-bit); • Apple Mac OS; Интерфейс пользователя также доступен из командной строки	
3.	Монитор(экран)	Отображение изображения	Экран с разрешением 1024x768px и более	Экран с разрешением от 1920×1080px.
4.	Платформа для разработки приложений	Среда разработки	JRE 17 – Java Runtime Environment версии 17	

4. Руководство администратора/пользователя

4.1. Установка

1. Необходимо скачать и распаковать архив.

Архив содержит:

- jar-приложение;
- файл `application.yml` конфигурации приложения;
- файл `init.sql`, который содержит инструкции по созданию и подготовке базы данных.

2. Запуск производится в командной строке в JRE 17 с помощью команды:

```
java -jar sms-analyzer.jar --spring-config-location=*путь до распакованной папки*\application.yml
```

3. По инструкциям из файла `init.sql` нужно создать базу данных postgres, содержащую процедуру вставки результатов и три таблицы:

- `message` — таблица с сообщениями;
- `template` — таблица с шаблонами;
- `result` — таблица с результатами.

4.2. Основные компоненты

1. Логи приложения: хранятся в папке `logs` рядом с jar-приложением.
2. Интерфейс взаимодействия с приложением: интерфейс базы данных (командная строка, pgAdmin или другие способы работы с базой данных).
3. Конфигурация (определяется в `application.yml`):

- Блок `datasource` — настройка подключения к базе данных:

```
datasource:  
  url: "jdbc:postgresql://localhost:5432/test?prepareThreshold=0"  
  username: test  
  password: test  
  hikari:  
    minimum-idle: 1
```

- Блок `matching-sms` — настройка времени, как часто осуществляется обработка сообщений, а также максимальный размер сообщений за один раз:

```
matching-sms:  
  schedule: '*/15 * * * * *'  
  size: 1000000
```

4.3. Результаты обработки шаблонов

Таблица с шаблонами(пример текста):

id	template
1	%d{1,3} Неудачная попытка платежа Причина отказа: %w{1,4}
2	Уважаемый клиент Ваша карта %d активирована
3	%d Счет %d Отмена платежа %d %w Баланс %d %w
4	%d Счет %d Отмена платежа %d руб Баланс %d руб
5	%d{1,2} (Мск) На Вашем счету недостаточно средств
6	%d{1,2} (Мск) %d Пополнение Вклада %d{1,2} руб Баланс %d руб

Таблица с сообщениями (колонка done — признак, что сообщение было обработано, может быть true/false):

id	message	done
1	10.01.25 10:00 (Мск) *0000 Пополнение Вклада 0000 1000000.00 руб Баланс 10000.00 руб	t
2	10.01.24 Счет 0000 Отмена платежа 500 руб Баланс 100.00 руб	t
3	10.01.24 Счет 0000 Отмена платежа 0000 Баланс 100.00	t
4	10.01.25 На Вашем счету недостаточно средств	t
5	10.01.25 10:00 (Мск) На Вашем счету недостаточно средств	t
6	10.01.25 10:00 Неудачная попытка платежа Причина отказа: недостаточно средств	t

Таблица результатов:

id_message	id_template
1	6
2	4
2	3
3	3
5	5
6	1

5. Алгоритм составления дерева поиска

Алгоритм работы:

- Сервис запрашивает из БД шаблоны и пакет СМС для анализа. Пакет СМС определяется по последней дате выгрузки из БД, т.е. выгружаются все СМС, дата которых более поздняя, чем дата выгрузки.
- В БД обновляется дата последней выгрузки СМС.
- Шаблоны разбиваются на части (далее Условия), разделителем является пробел:

%d{1, 2} Неудачная попытка %d	→	%d{1, 2} Неудачная	попытка	%d
-------------------------------	---	--------------------	---------	----

4. Из полученных Условий, Сервис строит дерево поиска.
5. Каждая СМС разбивается на части, разделителем является пробел, и анализируется по дереву поиска, в результате, каждой СМС ставится в соответствие один или несколько шаблонов, если они были найдены.
6. Результат соответствия каждой СМС с шаблонами записывается в БД в виде: id сообщения → id шаблона.
 - Если одной СМС соответствует один шаблон, то будет одна запись в БД.
 - Если одной СМС соответствует несколько шаблонов, то и записей в БД будет несколько.
 - Если одной СМС не соответствует ни один шаблон, то вместо id шаблона будет null.
7. Таким образом обрабатывается весь пакет СМС.

Шаблон может состоять из следующих «Условий»:

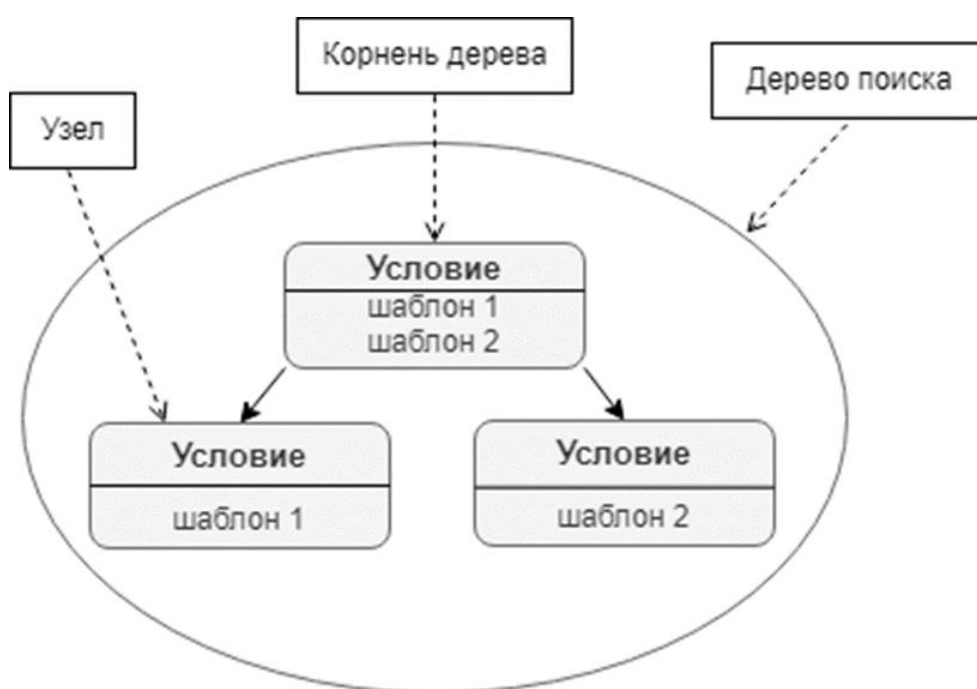
Условие	Описание	Дополнительно
константная строка	постоянный непрерывный набор букв, цифр и/или спецсимволов	
%w	любой непрерывный набор букв , цифр и/или спецсимволов	может быть пустым
%d	любой непрерывный набор цифр и/или спецсимволов	может быть пустым
%w{1, n}	ограниченная последовательность слов (состоящих из букв, цифр или спецсимволов), разделенных пробелом, где n – число слов (слов должно быть не менее 1 и не более чем n)	Значение n должно быть не более 20 ; %w{1,1} не использовать
%d{1, n}	ограниченная последовательность чисел (состоящих из цифр или спецсимволов), разделенных пробелом, где n – число чисел (чисел должно быть не менее 1 и не более чем n)	Значение n должно быть не более 20 ; %d{1,1} не использовать
%w+	последовательность слов (состоящих из букв, цифр или спецсимволов), разделенных пробелом или несколькими пробелами	

%d+	последовательность чисел (состоящих из цифр или спецсимволов), разделенных пробелом или несколькими пробелами	
------------	--	--

После выгрузки шаблонов из БД и разбиения их на «Условия» (см. схему), из этих частей образуется дерево поиска.

Дерево поиска состоит из узлов, первый узел дерева называется корнем дерева.

Каждый узел содержит «Условие» и список соответствующих условию шаблонов (в данном узле дерева поиска):



Пример сообщения и подходящих ему шаблонов:

01.01.2025 01:01 (МСК) Блокировка карты ***0123** до погашения задолженности. Для получения информации обратитесь в банк, тел. **88001234567**

%d+ (МСК) Блокировка карты **%w{1,10}** Для получения информации обратитесь в банк, тел. **%d**

%d{1, 2} (МСК) Блокировка карты **%w+** Для получения информации обратитесь в банк, тел. **%d**

Далее последовательно берется каждая часть шаблона для последующей обработки.

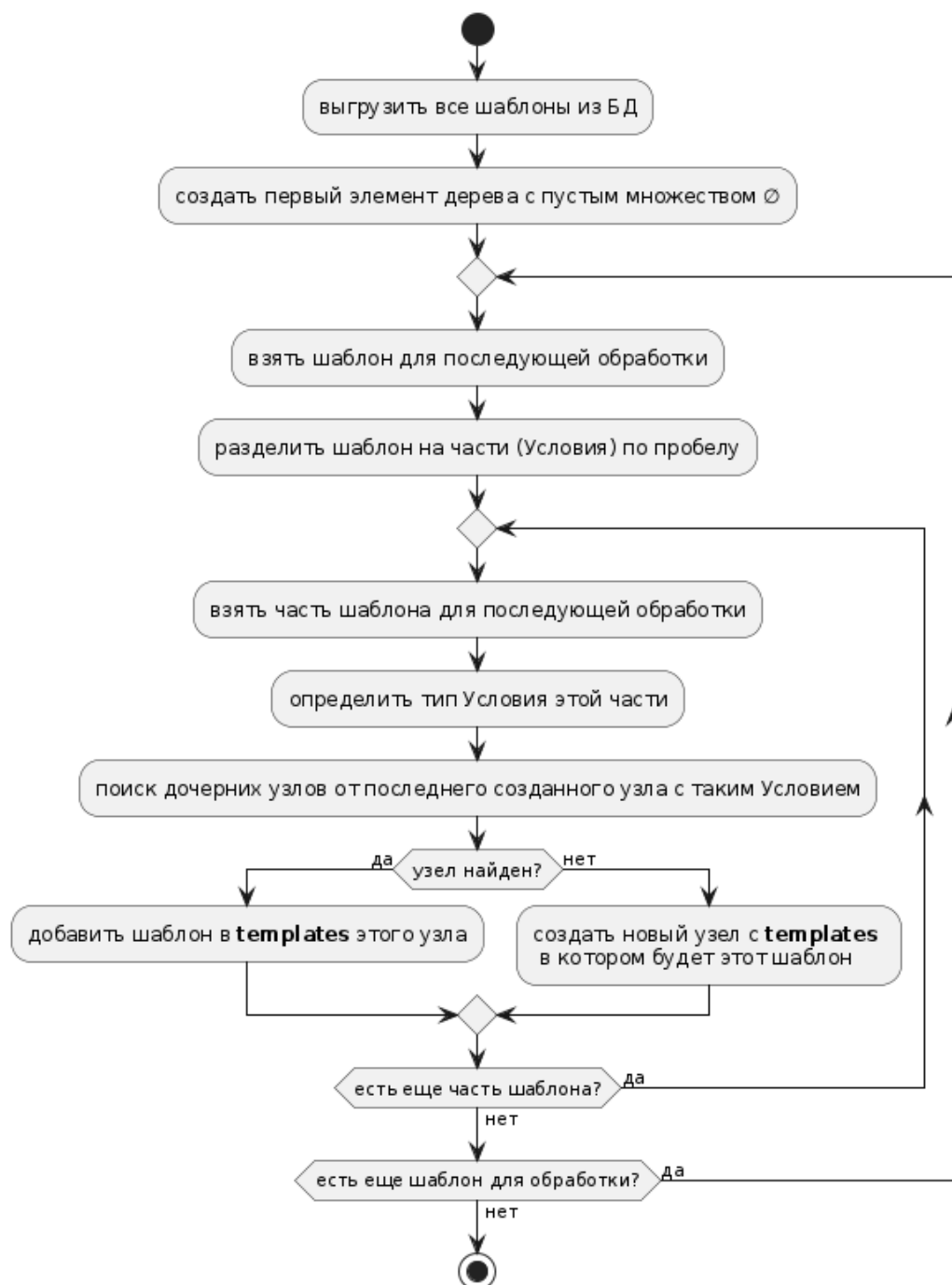
Определяется тип Условия данной части (описаны выше) и выполняется поиск существующего узла с таким условием.

Если узел найден, то шаблон добавляется в **templates** (лист шаблонов) этого узла.

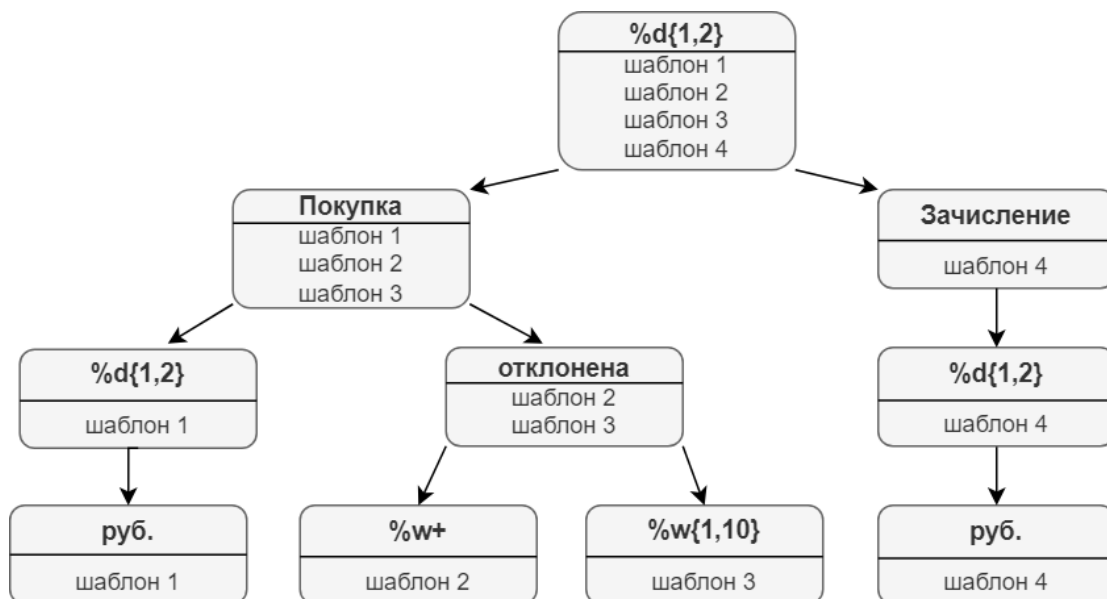
Если узел не найден, добавляется новый узел, содержащий текущий **conditions** (условие) и **templates** в котором будет один шаблон (который обрабатывается в данной итерации).

Таким образом последовательно обрабатываются все шаблоны.

Диаграмма активности "Построения дерева поиска"



Пример дерева для четырех шаблонов:



Название	Шаблон
Шаблон 1	%d{1,2} Покупка %d{1,2} руб.
Шаблон 2	%d{1,2} Покупка отклонена %w+
Шаблон 3	%d{1,2} Покупка отклонена %w{1,10}
Шаблон 4	%d{1,2} Зачисление %d{1,2} руб.

6. Ограничения

Программа может не справляться с большим количеством сообщений или сложными шаблонами.

Результаты анализа могут быть неточными из-за ошибок в шаблонах или других факторов.

7. Авторские права

Данное ПО защищено авторскими правами, запрещается копирование, модификация и распространение данного ПО без согласия правообладателя.