



Метод распределенного хранения аудио-файлов в NoSQL базе данных

Студент: Наместник Анастасия Андреевна

Группа: ИУ7-83Б

Руководитель: Гаврилова Юлия Михайловна

Москва
2022

Целью выпускной квалификационной работы является разработка метода распределенного хранения аудио-файлов в NoSQL базе данных

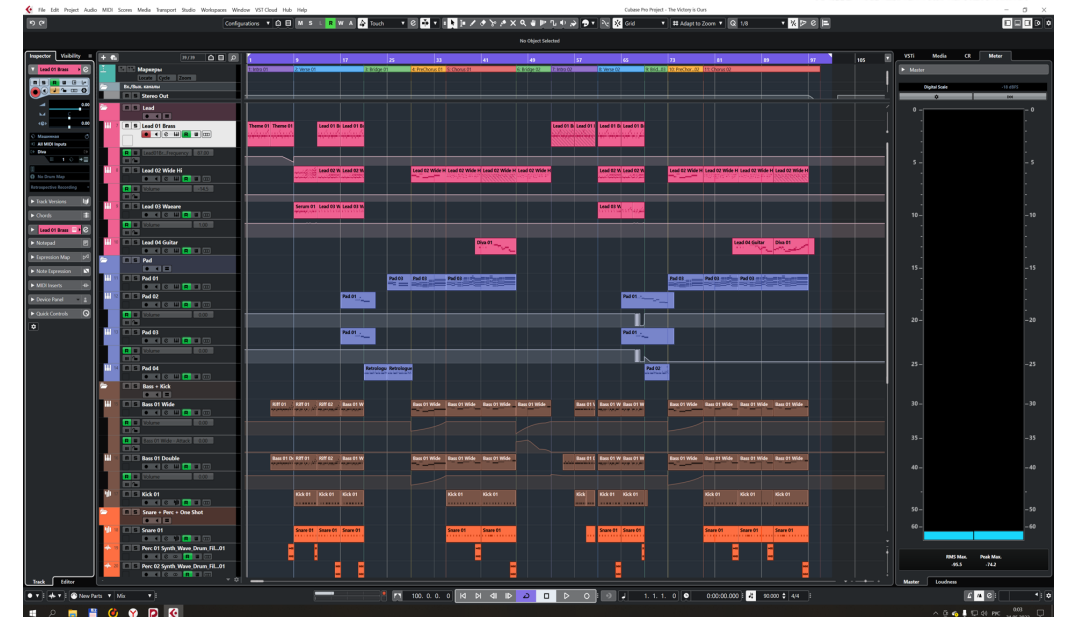
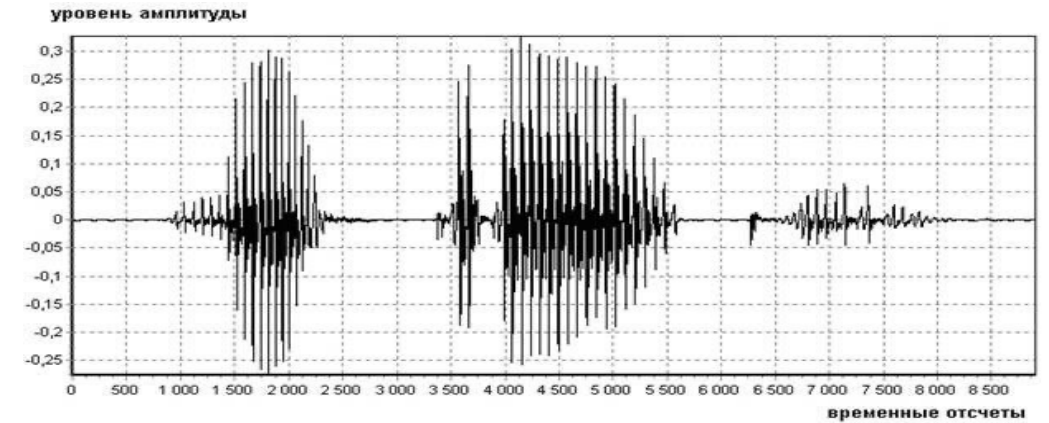
Задачи:

- проанализировать существующие аудио-форматы, рассмотреть способы представления и хранения звуковой информации;
- разработать метод хранения аудио-файлов в базе данных;
- реализовать разработанный метод;
- исследовать зависимость времени работы операций вставки и извлечения в базе данных от количества дорожек в аудио-файле.

Представление аудио-информации

Музыкальный аудио-файл может быть представлен:

- Оцифрованный звук
 - медиа-плееры и сопровождение видео ряда;
 - передача данных в сети;
 - профессиональная студийная запись.
- Цифровой интерфейс музыкальных инструментов
 - работа композитора;
 - звуковой дизайн приложений;
 - сведение музыкальных дорожек.



Анализ форматов аудио-файлов

Тип файла / Критерий	WAV	MP3	FLAC	MIDI
Тип хранимой информации	Оцифрованный звук	Оцифрованный звук	Оцифрованный звук	Цифровой интерфейс музыкальных инструментов
Тип сжатия	Без сжатия	Сжатие с потерями	Сжатие без потерь	Неприменимо
Область применения	Профессиональ- ная обработка аудио-звука	Хранение данных и передача по сети	Звуковое сопровождение к HD-видео, медиаплееры	Управление музыкальным оборудованием

Формат MIDI-файла

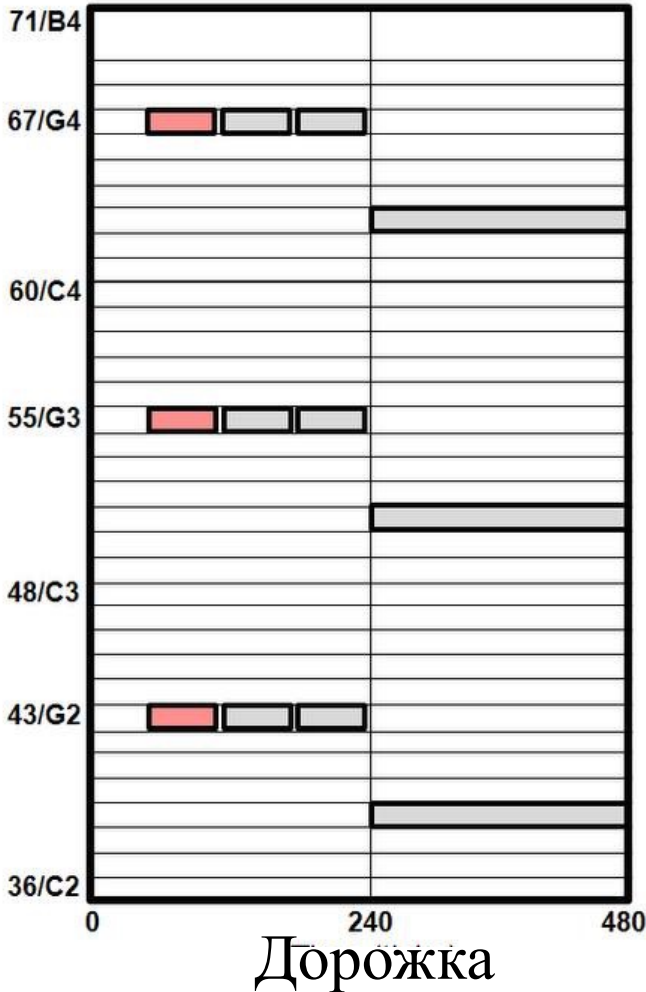
Дорожка – последовательность команд определенного инструмента для получения желаемого звука



Нотный стан

Время (тики)	Сообщение	Канал	Номер ноты	Громкость
60	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	63	100
0	NOTE ON	2	51	100
0	NOTE ON	2	39	100
240	NOTE OFF	1	63	0
0	NOTE OFF	2	51	0
0	NOTE OFF	2	39	0

Таблица MIDI-команд



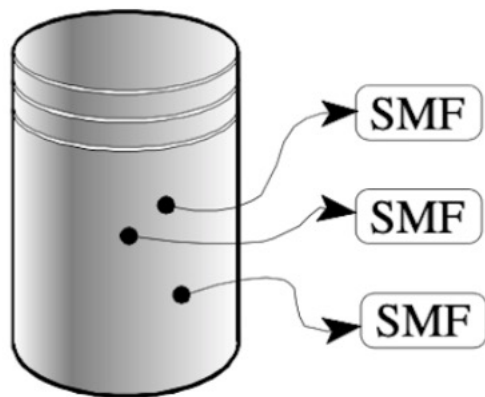
Дорожка

Спецификация MIDI-файла



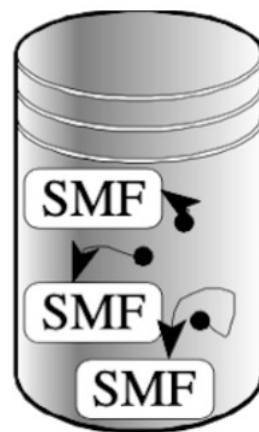
Способы хранения MIDI-файлов в базе данных

База данных путей к файлу



- + Компактный объем данных в БД
- Доступ к конкретному элементу данных
- Целостность данных

База данных массива байтов



- Доступ к конкретному элементу данных
- Целостность данных
- Компактный объем данных в БД

База данных структуры файла



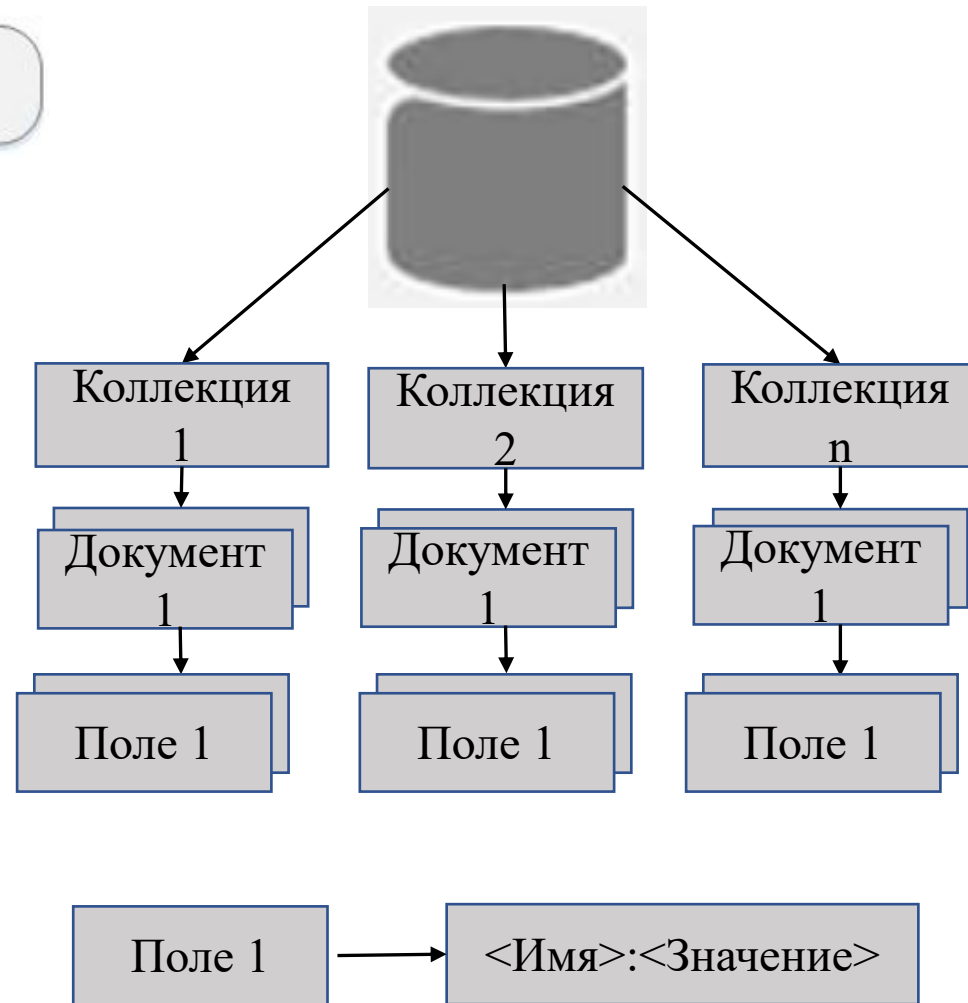
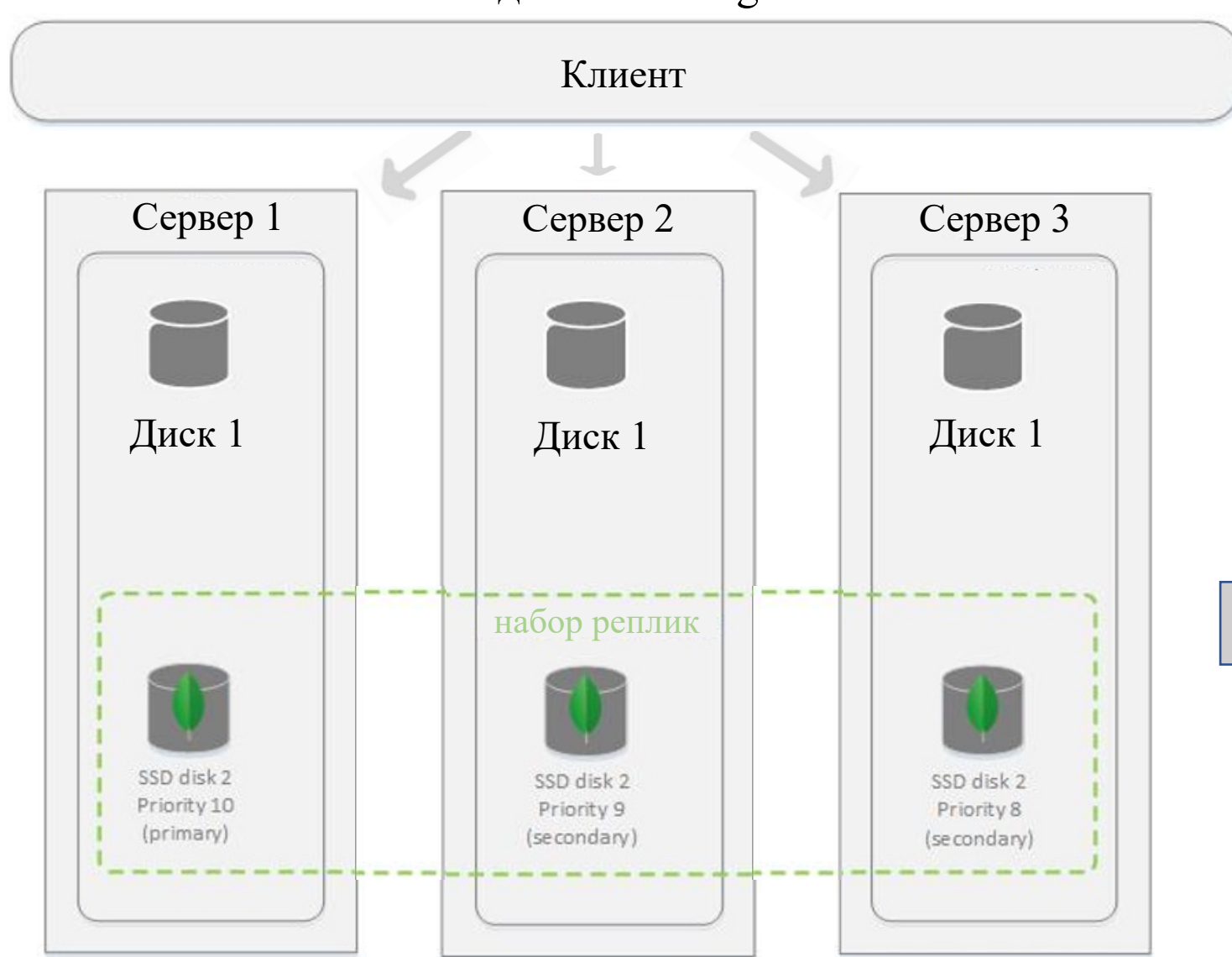
- + Доступ к конкретному элементу данных
- + Целостность данных
- Компактный объем данных в БД

Анализ моделей хранения данных на основе структуры

Модель / Критерий	Ключ-значение	Семейства столбцов	Графы	Документы
Связанность данных	Слабо связаны	Слабо связаны	Сильно связаны	Слабо связаны
Ограничения	Количество пар ключ-значение	Количество столбцов	Количество узлов	<ul style="list-style-type: none"> Размер документа Уровень вложенности
Оптимизация по типу данных	—	+	—	—
Пример СУБД	<ul style="list-style-type: none"> Riak Dynamo Redis 	<ul style="list-style-type: none"> HBase Cassandra 	<ul style="list-style-type: none"> InfoGrid Infinite Graph 	<ul style="list-style-type: none"> CouchDB MongoDB

Клиент-серверная архитектура MongoDB

База данных MongoDB



Функциональная диаграмма записи MIDI-файла в MongoDB

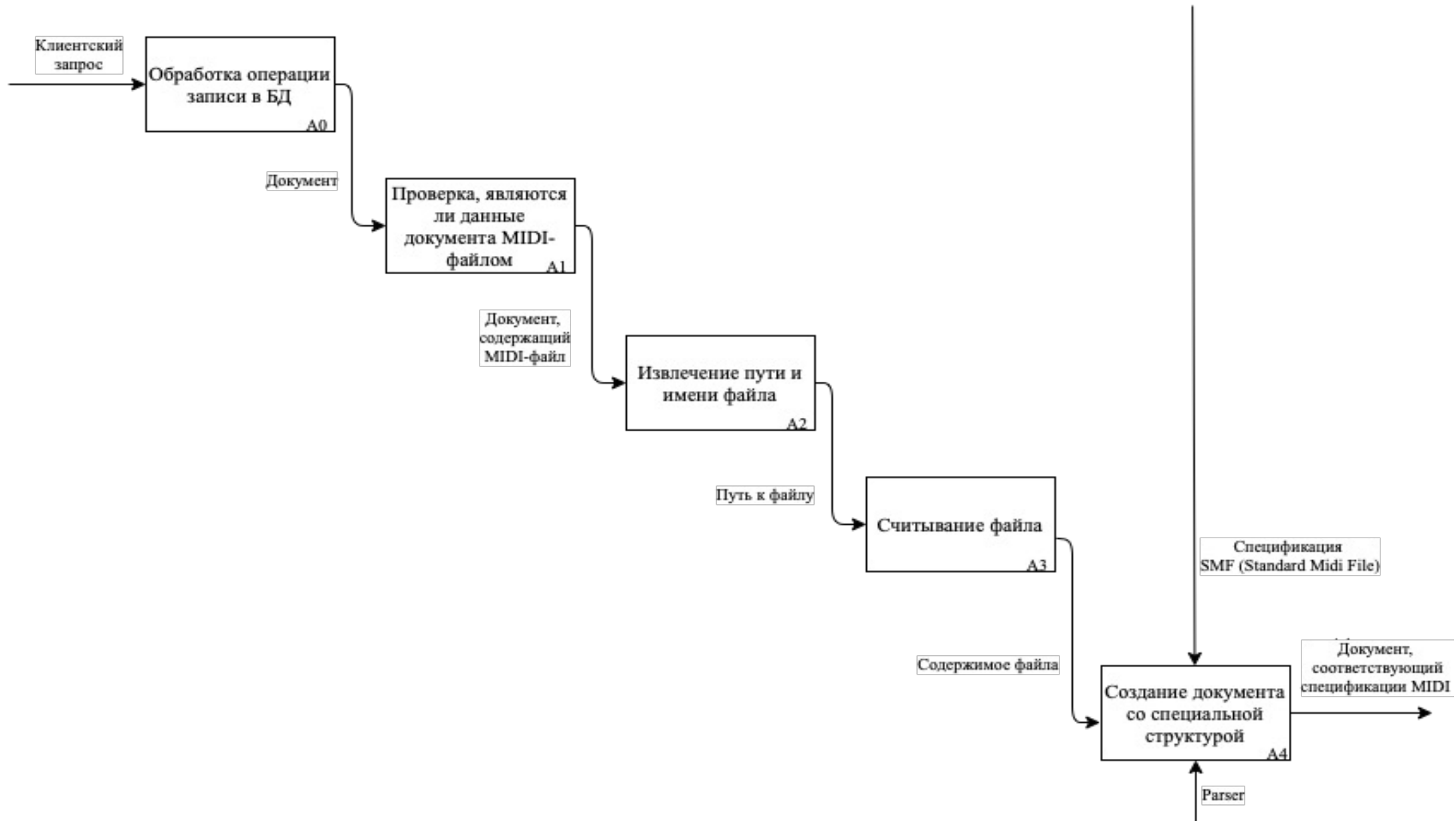
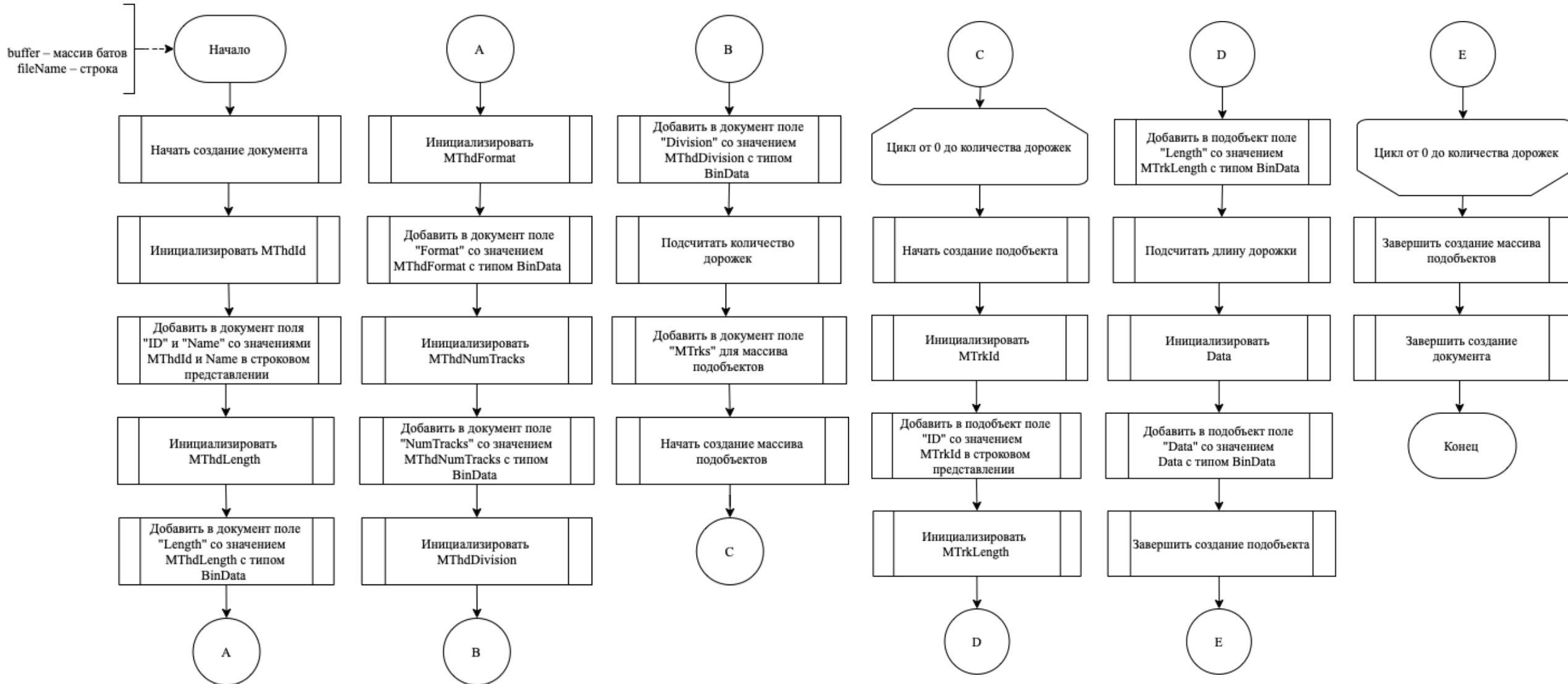


Схема создания документа со структурой MIDI-файла



Функциональная диаграмма чтения MIDI-файла из MongoDB

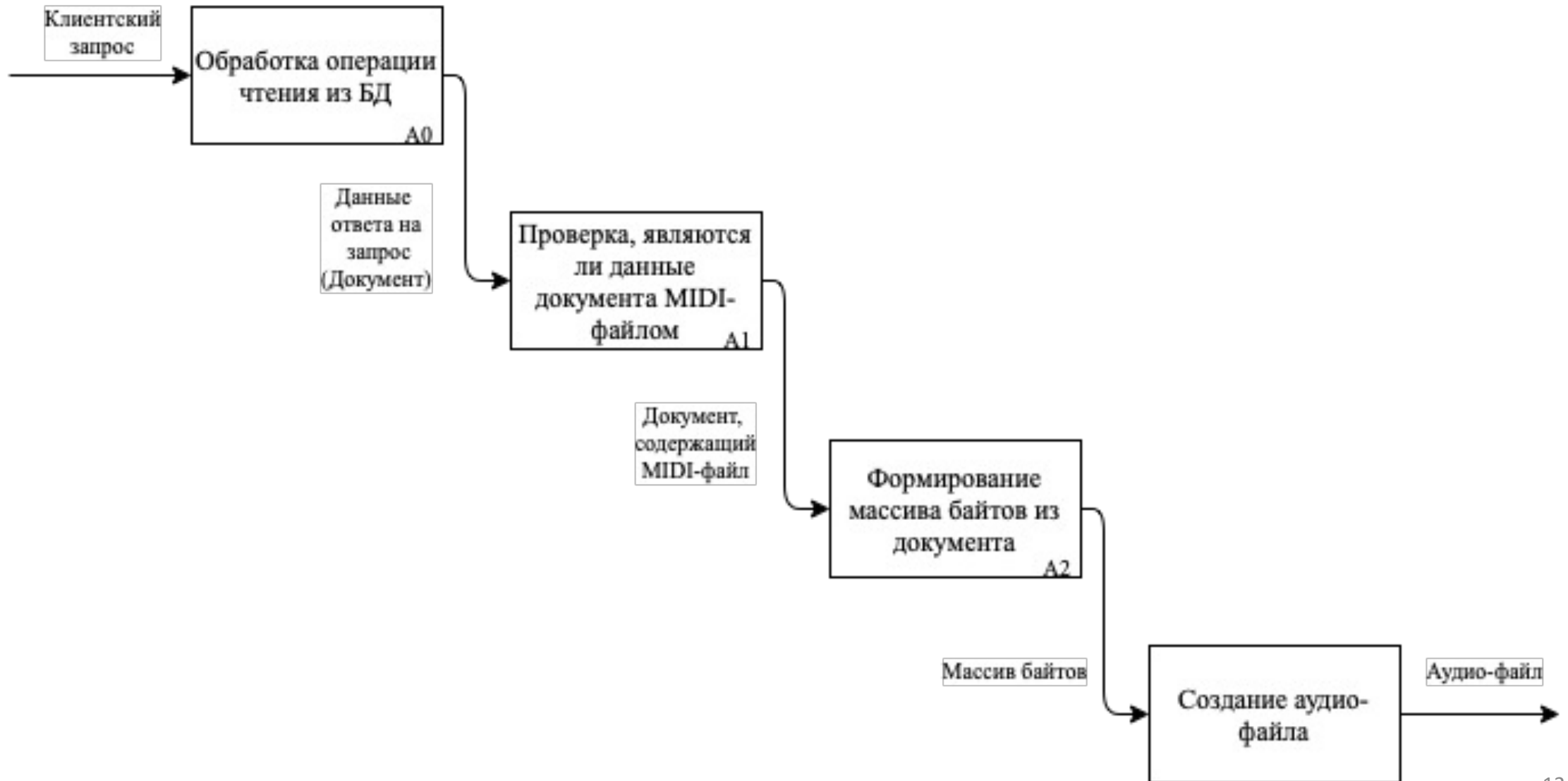


Схема создания MIDI-файла из документа MongoDB

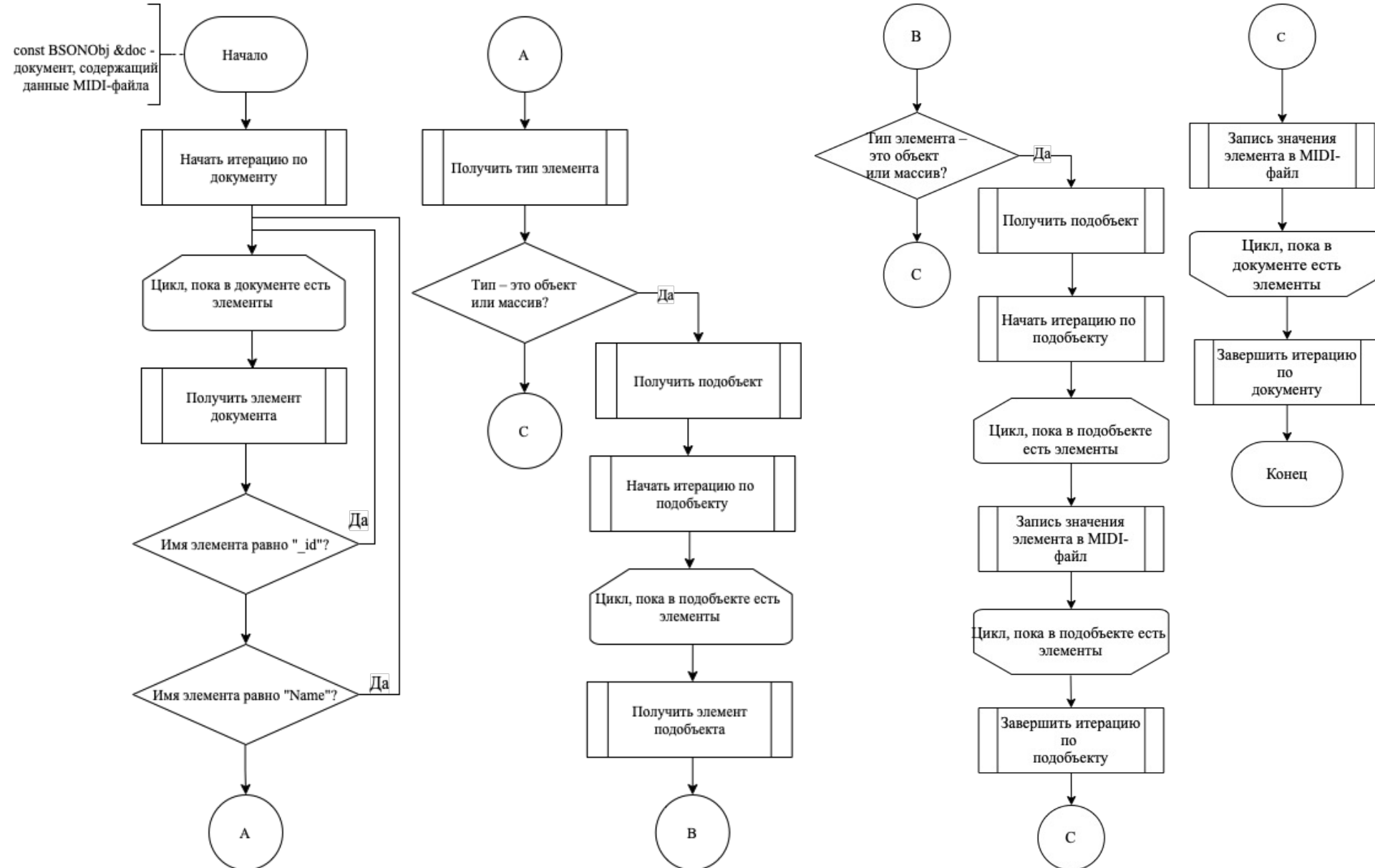
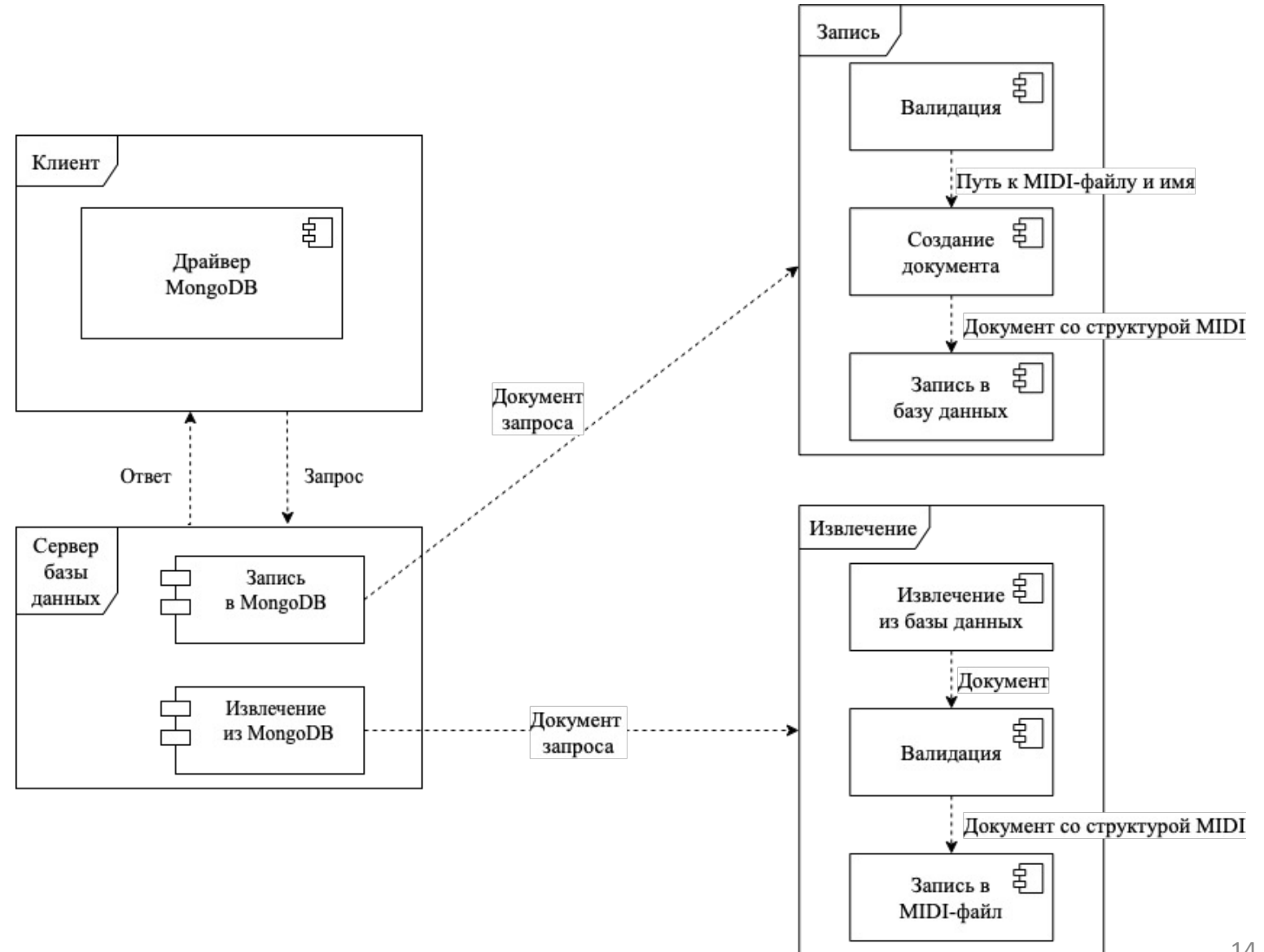
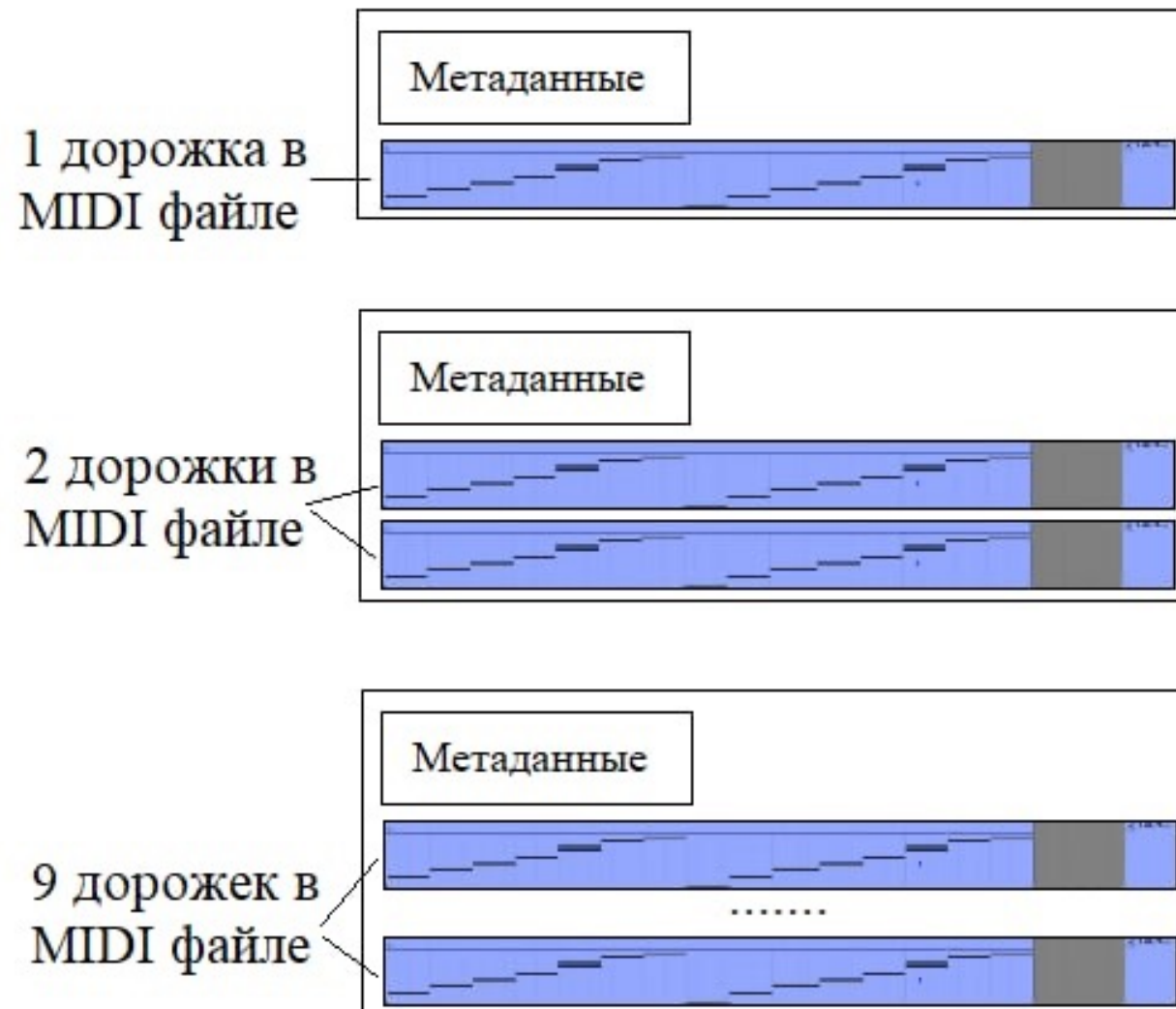


Схема разработанного программного обеспечения

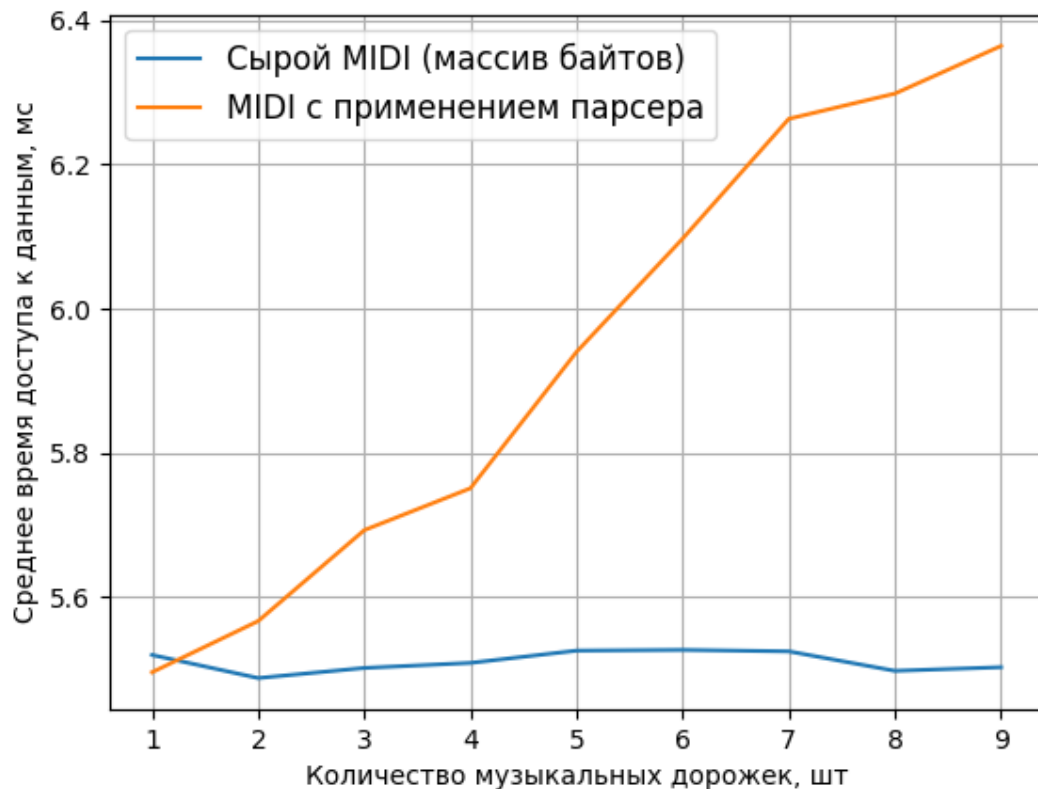
- Разработано:
 - Валидация при записи и чтении
 - Создание документа
 - Запись в MIDI-файл
- Доработано:
 - Запись в MongoDB
 - Извлечение из MongoDB



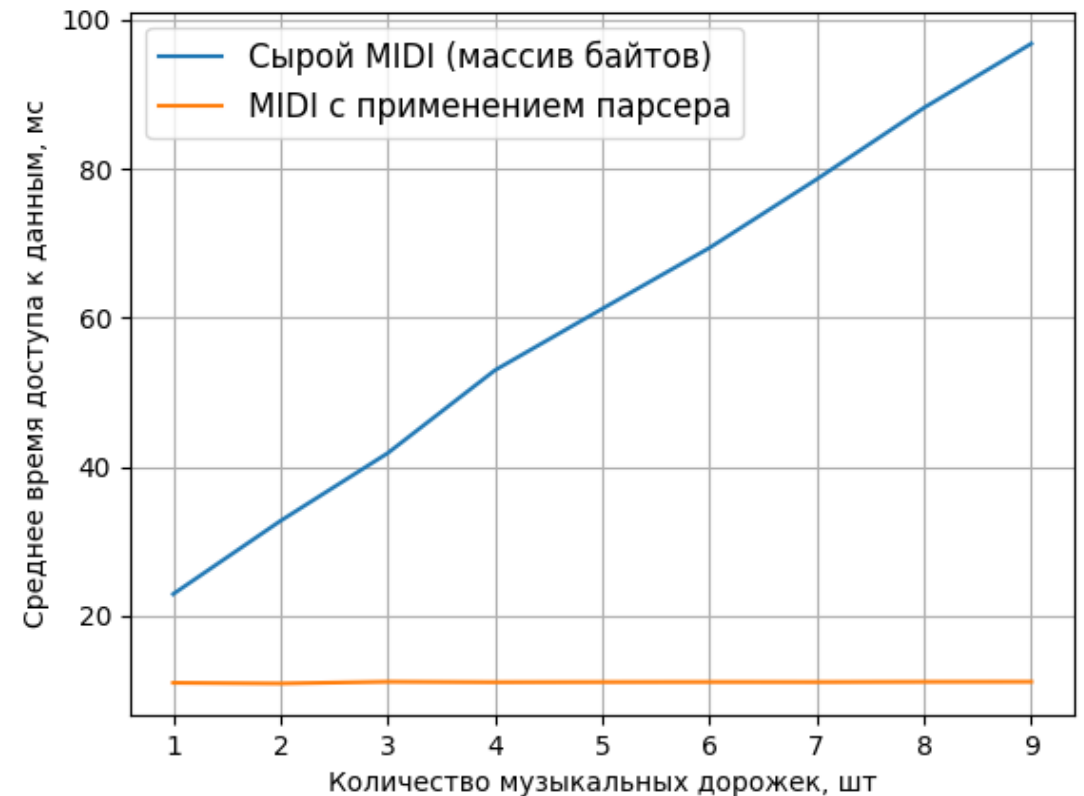
Подготовка данных для исследования



Зависимость времени работы операций вставки и чтения в базе данных от количества дорожек в аудио-файле.

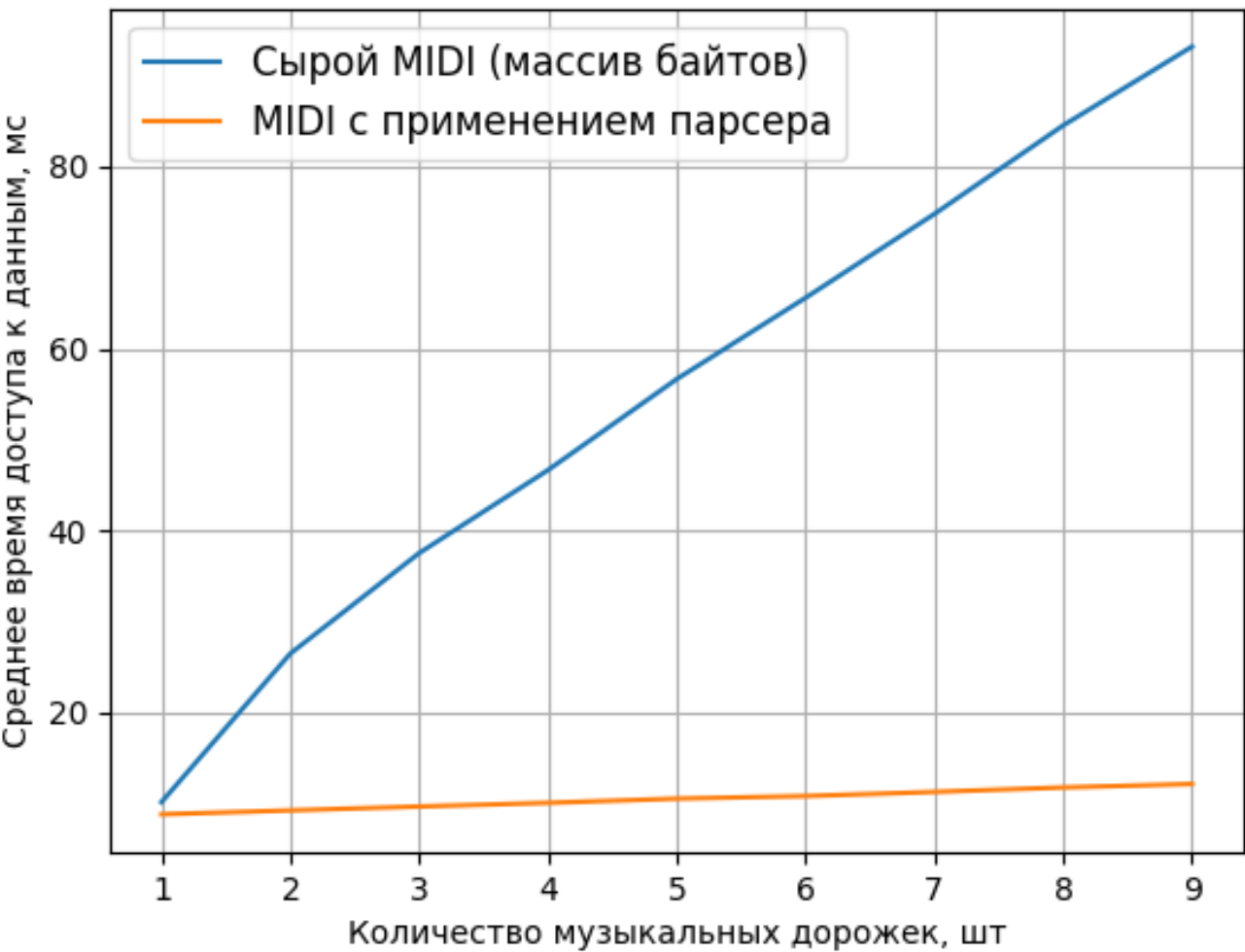


Операция записи MIDI-файла в базу данных



Операция чтения MIDI-файла из базы данных

Зависимость времени работы последовательных операций вставки и чтения из базы данных от количества дорожек в аудио-файле



Скорость роста (мс / шт)	
Метод хранения на основе структуры	Метод хранения на основе массива байтов
≈ 0,37	≈ 9,22

Заключение

Поставленная цель была достигнута. Были решены следующие поставленные задачи:

- проанализированы существующие аудио-форматы, рассмотрены способы представления и хранения звуковой информации;
- разработан метод хранения аудио-файлов в базе данных;
- реализован разработанный метод;
- исследована зависимость времени работы операций вставки и удаления в базе данных от количества дорожек в аудио-файле.

Направление дальнейшего развития

- добавить пользователю возможность задавать тип данных, в котором будут храниться поля аудио-файла в базе данных;
- разработать гибкую структуру API записи/чтения аудио-файла в/из базы данных.