

УДК 004.832.22

АЛГОРИТМ РАБОТЫ СИСТЕМЫ АВТОМАТИЧЕСКОЙ АУДИОДЕСКРИПЦИИ

Воробьев Ярослав Вадимович, аспирант, Базовая кафедра «Аналитика больших данных и методы видеонализа», Уральский федеральный университет

Аннотация

Любая информационная система работает в соответствии со своими внутренними алгоритмами. В случае системы на основе искусственного интеллекта, разработка подробного алгоритма становится ещё более важной задачей: необходимо заранее сократить количество ошибок системы, просчитав возможные сценарии её работы. Так для системы автоматической аудиодескрипции было решено спроектировать алгоритм, который в дальнейшем будет использоваться непосредственно при разработке такой системы. В статье представлена упрощённая схема алгоритма, получившаяся на первых этапах исследования, а также декомпозиция его блоков и подробное описание таблиц, к которым данная система обращается при прохождении шагов алгоритма. Предполагается, что к последнему пункту алгоритма в базе данных будут собраны все необходимые для озвучки видеоряда текстовые описания и соответствующие им временные интервалы.

КЛЮЧЕВЫЕ СЛОВА: алгоритм, автоматическая аудиодескрипция, нарушения зрения, визуализация базы данных.

OPERATING ALGORITHM OF THE AUTOMATIC AUDIO DESCRIPTION SYSTEM

Vorobyov Yaroslav Vadimovich, postgraduate, Dept. of Big Data Analytics and Video Analysis Methods, Ural Federal University

Abstract

Any information system operates in accordance with its internal algorithms. In the case of a system based on artificial intelligence, the development of a detailed algorithm becomes an even more important task: it is necessary to reduce the number of system errors in advance by calculating possible scenarios for its operation. So, for an automatic audio description system, it was decided to design an algorithm that will subsequently be used directly in the development of such a system. The article presents a detailed algorithm and description of the tables that this system accesses when going through the steps of the algorithm. It is assumed that by the last point of the algorithm, all the text descriptions necessary for dubbing the video sequence and the corresponding time intervals will be collected in the database.

KEYWORDS: algorithm, automatic audio description, visual impairment, database visualization.

1. Введение

Актуальность данной работы обусловлена необходимостью разработки алгоритма для разрабатываемой системы автоматической аудиодескрипции на основе искусственного интеллекта (временное название – Ozari). Как и любая другая информационная система, система Ozari требует для своей работы чёткой и гибкой инструкции по последовательному выполнению действий – то есть, алгоритма. Проектирование алгоритма перед непосредственной реализацией позволяет в виде наглядных схем просчитать возможные события в системе и сценарии её поведения – в зависимости от типа данных, полученных на

входе, и их дальнейших преобразований. Кроме того, разработка алгоритма позволила уточнить структуру и предполагаемое содержание базы данных.

Таким образом, целью данной работы является построение алгоритма обработки фильмов системой автоматической аудиодескрипции. Для этого необходимо выполнить следующие задачи:

- выявить ключевую информацию, необходимую для обработки фильмов, на основе того, какие задачи выполняются тифлокомментаторами [1,2] в соответствии с установленным государственным стандартом [3];
- чётко определить и записать основные понятия, необходимые для описания алгоритма и разработки базы данных;
- составить схему базы данных, учитывая потребности алгоритма по получению, преобразованию и хранению информации;
- используя собранную информацию, составить алгоритм работы системы, декомпозируя при этом сложные этапы алгоритма на более простые шаги.

2. Материалы и методы

Материалам для разработки алгоритма послужила информация, полученная на прошлых этапах исследования данной масштабной работы. Так, были учтены уставленные ранее требования к разработке системы: продумана схема по определению типа интервала, что, в свою очередь, помогает системе выявлять ключевую информацию для каждой сцены. Кроме того предполагается, что ручное или полу-автоматическое внесение метаданных о фильмах в БД позволит системе совершать меньше ошибок, а сложность необходимого первичного обучения будет снижена, что обеспечит также требование по снижению сложности разрабатываемой системы.

В первую очередь при разработке алгоритма были выявлены его ключевые этапы. Без декомпозиции этих этапов на отдельные шаги, схема выглядела достаточно простой и занимала менее 10 шагов от получения данных до возврата информации в готовом виде. Декомпозиция позволила сделать схему гораздо более подробной, увеличив количество известных шагов более чем в 2 раза.

Для построения базы данных были использованы знания, полученные на первой ступени обучения в университете, а также личные профессиональные навыки. Для составления схемы базы данных был использован инструмент для визуального проектирования баз данных в MySQL Workbench [4].

Для дальнейшей работы были установлены некоторые термины:

- *сессия* – процесс обработки и создания аудиодескрипции для одного фильма;
- *интервал* – временной промежуток в фильме, когда герои на экране ничего не произносят, а также отсутствуют важные для смысла фильма звуковые эффекты или музыка. Имеет начальную и конечную точки, выраженные в формате чч:мм:сс.млс, и описание, соответствующее происходящему на экране между начальной и конечной временными точками;
- *сущность* – это любой объект, показываемый на экране в момент интервала: человек, животное, автомобиль, элемент природы, помещения – то есть, любая материя в пространстве. Сущность может иметь событие (действие), и чаще всего оно имеет свойства;
- *событие (действие)* – любое значимое (в рамках интервала) действие сущности, которое возможно описать глаголом.

3. Результаты и обсуждение

Как было отмечено в прошлых работах, система автоматической генерации аудиокomentarиев состоит из двух частей. Внешняя часть представляет из себя вебсайт в сети Интернет, на котором происходит получение, публикация и обработка заказов на создание тифлокомментариев. Внутренняя часть отвечает непосредственно за саму обработку видеоматериалов. Ранее была рассмотрена их взаимосвязь, в том числе, построена последовательная схема бизнес-процесса создания аудиодескрипции – от получения заказа и до выпуска готового материала. В данной же работе целью было исследовать более детально именно работу внутренней части системы, и на основании полученных данных спроектировать алгоритм (рисунок 1) [5-7].

Как видно по схеме алгоритма, он состоит из 10 элементов (учитывая начальный и последний шаг). При этом 3 элемента (STR, DEC и INF) на данной схеме представлены в сокращённом виде, и каждый из них будет подробнее рассмотрен далее.

Алгоритм описывает действия, выполняемые системой, при получении команды обработки фильма. При этом, для корректной работы искусственный интеллект системы должен уметь запоминать информацию (контекст) в рамках одной сессии – то есть, в течение обработки одного фильма. Поэтому первым пунктом алгоритма является обращение системы к метаданным из таблиц «Фильмы» (табл. 1) и «Герои» (табл. 2).

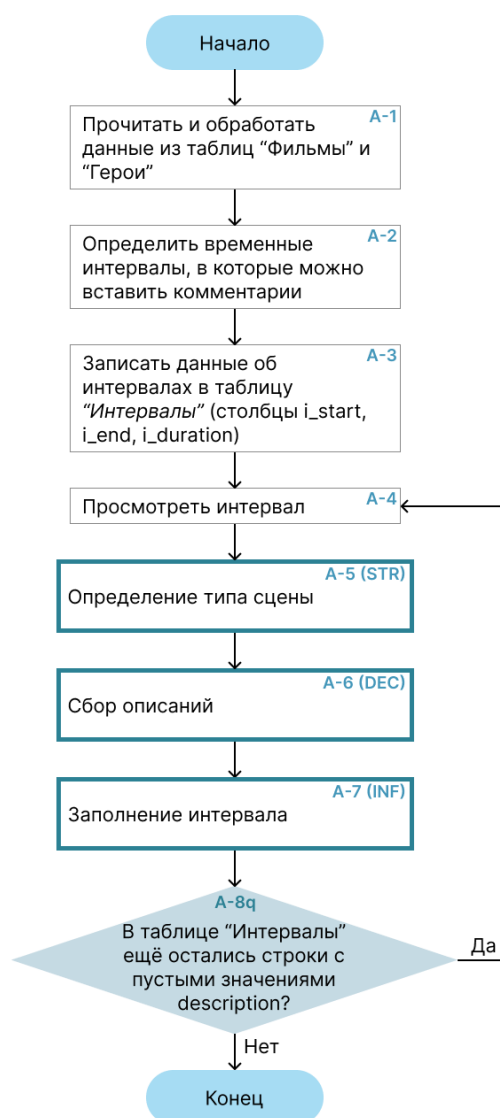


Рис. 1. Краткий вид алгоритма системы

Использование метаданных фильмов облегчает работу системы – так, ей изначально будет дана некоторая вводная информация о фильме, что, вероятно, позволит ей лучше понимать важность тех или иных событий ленты.

Таблица 1. Пример таблицы «Фильмы» (сокращённая версия)

Название	Год выхода	Жанр	Описание	Примечание
Крёстный отец	1972	Драма, криминал	Криминальная сага, повествующая о нью-йоркской сицилийской мафиозной семье Корлеоне. Фильм охватывает период 1945-1955 годов. Глава семьи, Дон Вито Корлеоне, выдаёт замуж свою дочь. В это время со Второй мировой войны возвращается его любимый сын Майкл. Майкл, герой войны, гордость семьи, не выражает желания заняться жестоким семейным бизнесом. Дон Корлеоне ведёт дела по старым правилам, но наступают	

Название	Год выхода	Жанр	Описание	Примечание
			иные времена, и появляются люди, желающие изменить сложившиеся порядки. На Дона Корлеоне совершается покушение.	
В погоне за счастьем	2006	Драма, биография	Крис Гарднер – отец-одиночка. Воспитывая пятилетнего сына, Крис изо всех сил старается сделать так, чтобы ребенок рос счастливым. Работая продавцом, он не может оплатить квартиру, и их выселяют. Оказавшись на улице, но не желая сдаваться, отец устраивается стажером в брокерскую компанию, рассчитывая получить должность специалиста. Только на протяжении стажировки он не будет получать никаких денег, а стажировка длится 6 месяцев...	
Тьма	2017	Триллер, фантастика, драма, криминал, детектив	История четырёх семей, живущих спокойной и размеренной жизнью в маленьком немецком городке. Видимая идиллия рушится, когда бесследно исчезают двое детей и воскресают тёмные тайны прошлого.	Путешествия во времени

Таблица будет заполняться или администратором системы, или самой системой автоматически при получении заказа на фильм. Столбец «Описание» содержит ключевую информацию о фильме – та, которая даётся самими авторами фильма или составляется редакторами онлайн-кинотеатров. Столбец «Примечание» указывает на дополнительную информацию о фильме – ту, которой нет в основном описании, но которая также может помочь системе при распознавании событий и описании сцен.

Такой же смысл – помощь системе в распознавании контекста – имеет таблица 2 («Герои») [8].

Таблица 2. Пример таблицы «Герои» (сокращённая версия)

Имя героя	Имя актёра	Главный герой	Описание	Примечание
Йонас Канвальд	Луис Хофманн	Да	Молодой (подросток), светлые волосы до плеч, голубые глаза, нормальный рост, нормальное телосложение	Школьник, эмоциональный
Марта Нильсен	Лиза Викари	Да	Молодая (подросток), тёмные волосы длинные волосы, тёмные глаза, нормальный рост, нормальное телосложение	Школьница
Ульрих Нильсен	Оливер Масукки	Нет	Взрослый, русые волосы средней длины, серые глаза, высокий, спортивное телосложение	Полицейский
Александр Кохлер	Питер Бенедикт	Нет	Взрослый, тёмные короткие волосы, серые глаза, высокий, нормальное телосложение	Директор станции

Предполагается, что, имея эту информацию, система быстрее и точнее будет определять героев на экране и приоритет их описания – в соответствии со столбцом «Главный герой». При этом таблица должна иметь полный список действующих лиц

произведения, кроме актёров массовки. Поскольку действия последних не оказывают влияние на ход сюжета, система опускает их описание.

На шаге А-2 система «отсматривает» фильм и определяет промежутки, в которые возможно вставить аудиокomentarии, и записывает соответствующую информацию в таблицу «Интервалы» (табл. 3). В дальнейшем именно эта таблица будет являться главным результатом работы системы – подготовленные данные о длительности интервала и его описание будут передаваться внешней части системы для отображения в пользовательском интерфейсе (в данном случае – в интерфейсе для актёров озвучки).

Привлекательным вариантом представляется автоматическое определение таких интервалов лишь по виду аудиодорожки – моментам наименьших звуковых колебаний – однако, помимо непосредственно голосов героев, колебания также могут вызываться шумами (звуками города, шагами и т.д.) и музыкой. Вероятно, существуют технологии отличия таких шумов и музыки от голосов, однако на данном этапе разработки системы предполагается полный «отсмотр» фильма системой.

Таблица 3. Пример таблицы «Интервалы» (сокращённая версия)

Тип	Начало	Конец	Доступное время	Описание
1	00:00:16.322	00:00:21.016	00:00:04.694	Показывается зелёный мрачный лес
2	00:00:44.058	00:00:50.301	00:00:06.243	На загородной дороге взрывается белый внедорожник
3	00:01:02.551	00:01:09.466	00:00:06.915	Напуганные молодые парень и девушка смотрят по сторонам
4	00:04:28.149	00:04:34.006	00:00:05.857	Эти парень и девушка смотрят на другого молодого человека, он спокоен, он показывает им на дом
4	00:08:13.173	00:08:19.324	00:00:06.151	Парень и девушка заходят в дом
4	00:08:23.226	00:08:27.063	00:00:03.837	Тёмноволосый мужчина средних лет встречает их

Результатом шагов А-2 и А-3 являются заполненные столбцы «Начало», «Конец» и «Доступное время» таблицы «Интервалы». Когда соответствующая информация получена и записана, с шага А-4 начинается главный цикл – просмотр и заполнение интервалов. Он будет повторяться, пока не будут заполнен столбец «Описание» таблицы «Интервалы» для всех строк.

Составление описаний начинается с блока А-5 (собственное имя STR – Scene Type Recognition, Распознавание Типа Сцены). Далее (рисунок 2) он представлен в развёрнутом виде.

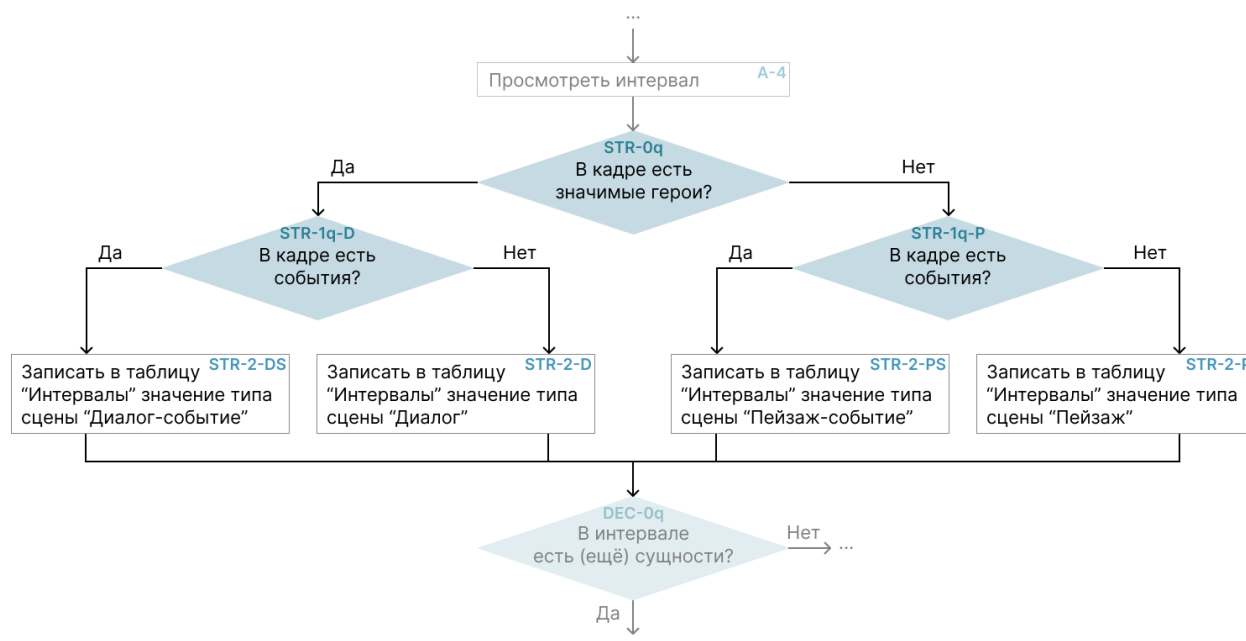


Рис. 2. Блок алгоритма – «STR»

Результатом данного шага будет дополнение таблицы «Интервалы» информацией о типе интервала (типе сцены). Знание типа сцены в дальнейшем позволяет системе определять, что необходимо описывать в текущем интервале. Как видно по схеме, всего есть 4 типа сцены:

- Диалог. На экране не происходит никаких значимых действий, кроме диалога. В моменты, когда герои молчат, следует описывать их эмоции, поведение, направление взора. Если описывать нечего, следует описывать обстановку – саму сцену.
- Пейзаж. На экране не происходит никаких значимых действий, лишь показывается некоторое пространство. Следует описывать это пространство – свет, мебель, общий вид помещения, если это помещение, или же, в противном случае, детали природы или города.
- Диалог-событие. Требования такие же, как к обычному Диалогу, однако приоритет для описания получает некоторое событие – поцелуй героев, их взятие за руки, или же, наоборот, борьба.
- Пейзаж-событие. Требования такие же, как к обычному Пейзажу, однако приоритет для описания получает некоторое событие – взрыв автомобиля, открытие дверей автобуса, появление летающей тарелки.

Таким образом, пейзажем являются все сцены, в которых не задействованы значимые герои; диалогом – наоборот, те сцены, в которых они задействованы. Имеющееся событие

понижает стандартные приоритеты для обоих типов сцен, и выставляет значение приоритета «Событие» равным 3.

Следующий шаг алгоритма – блок А-6 (собственное имя DEC – Descriptions Collecting, Сбор Описаний) представлен на рисунке 3.

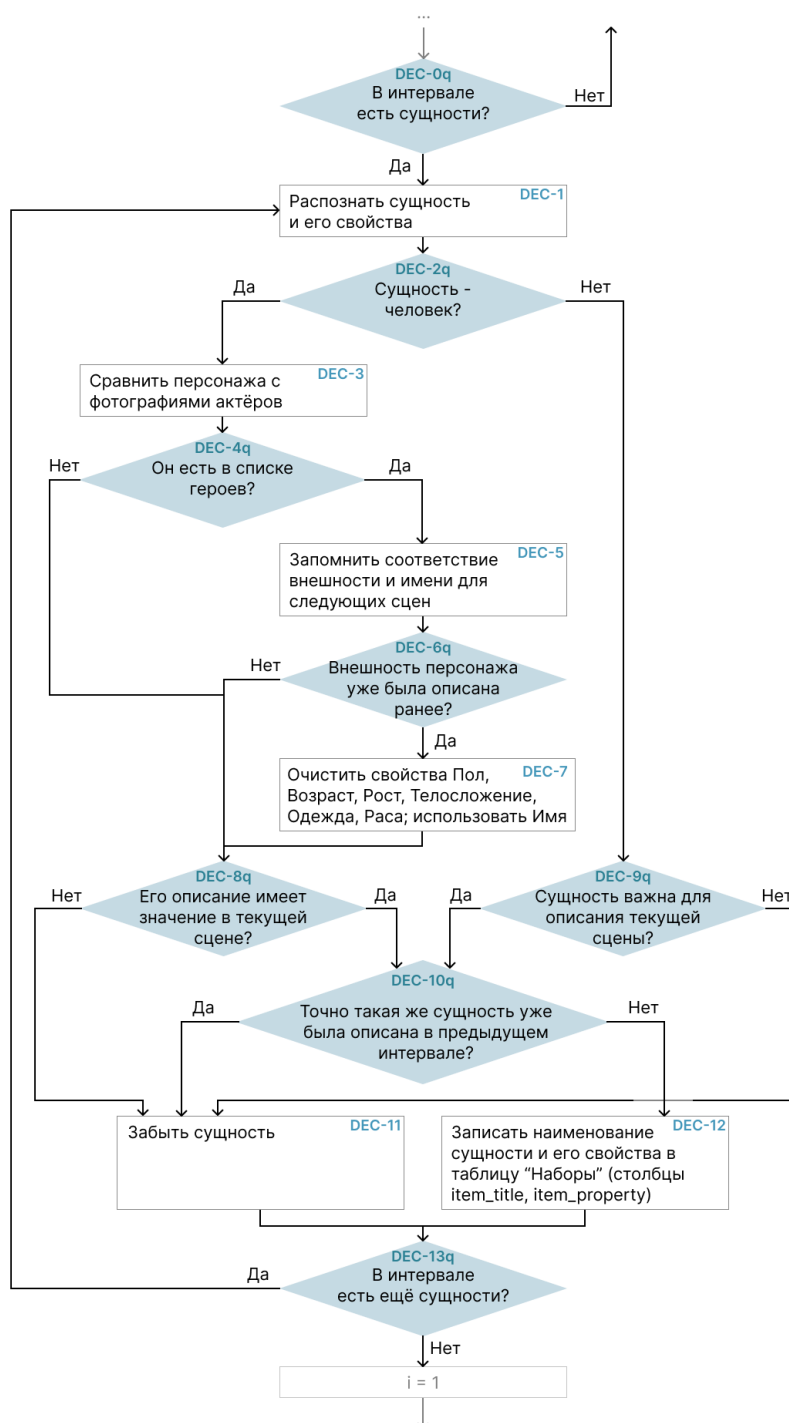


Рис. 3. Блок алгоритма – «DEC»

На данном шаге система выявляет сущности – то есть, то, что возможно описать. Сущность – это объект: человек, животное, автомобиль, элемент природы, помещения – то есть, любая материя в пространстве. Сущность может иметь событие (действие), и чаще всего оно имеет свойства, примеры которых можно увидеть в следующей таблице (табл. 4).

Таблица 4. Пример таблицы «Сущности» (сокращённая версия)

ID персонажа	Сущность	Событие	Качества
-	Лес	-	Зелёный, много деревьев, густой, высокий
-	Небо	-	Ясное, солнечное, голубое
-	Дорога	-	Загородная, узкая
-	Машина	Взрывается	Белая, внедорожник
1	Йонас Канвальд	Смотрит по сторонам	Мужчина, молодой, высокий, голубые глаза, русые волосы, встревоженный
2	Марта Нильсен	Смотрит по сторонам	Женщина, молодая, обычный рост, карие глаза, русые волосы, встревоженная
1	Йонас Канвальд	Смотрит на персонаж ID = 3	Встревоженный, напуганный
2	Марта Нильсен	Смотрит на персонаж ID = 3	Встревоженная, напуганная
3	Ульрих Нильсен	Показывает на дом	Мужчина, молодой, спокойный
1	Йонас Канвальд	Заходит в дом	-
2	Марта Нильсен	Заходит в дом	-
4	Александр Кохлер	Встречает персонаж ID = 1 и Встречает персонаж ID = 2	Мужчина, среднего возраста, высокий, карие глаза, тёмные волосы

Как видно по таблице, на данном этапе алгоритма в базе уже имеется набор информации об объектах – их наименования, действия и качества, при этом, первично эта информация была признана имеющей значение. Однако по-прежнему это только набор несогласованных по правилам русского языка слов. Для приведения этих наборов к ожидаемому виду, выполняется последний блок алгоритма – А-7 (собственное имя INF – Interval Filling, Заполнение Интервала), представленный на рисунке 4.

В заключительном блоке, как видно на схеме, имеется цикл. Цель этого цикла – добиться соответствия длины описания и имеющегося временного интервала для озвучивания этого описания. Для этого сначала к свойствам сущностей подбираются возможные синонимы, а если возможные синонимы закончились или их не было изначально – идёт сокращение в соответствии с таблицей приоритетов (табл. 5).

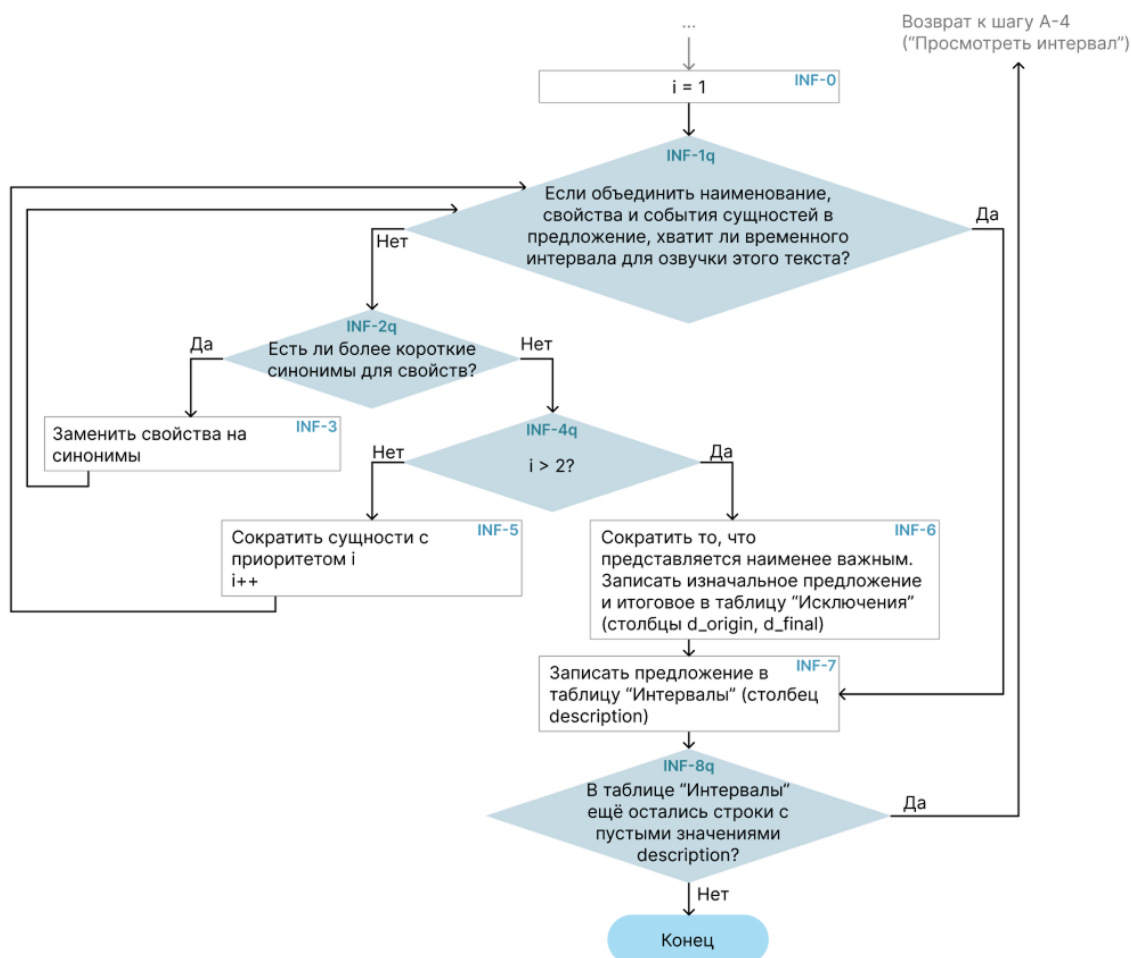


Рис. 4. Блок алгоритма – «INF»

Таблица 5. Пример таблицы «Приоритеты» (сокращённая версия)

Тип сцены	Сцена	Событие	Эмоции	Пол	Возраст	Рост	Телосложение	Одежда
Пейзаж	3	0	0	0	0	0	0	0
Пейзаж-событие	3	3	0	0	0	0	0	0
Диалог	2	0	3	3	3	2	2	1
Диалог-событие	1	3	3	3	3	2	2	1

Приведённая выше таблица показывает системе приоритеты того, что необходимо описывать в первую очередь, в зависимости от ранее определённого типа сцены. Так, приоритет 0 – сущность не описывается вовсе, 3 – сущность описывается обязательно. Например, для интервалов с типом сцены «Пейзаж» система не должна описывать события (потому что, исходя из типа сцены, в ней нет событий), а также все свойства, связанные с людьми (поскольку, также в соответствии с типом сцены, в ней нет людей). А для типа сцены «Диалог» множество сущностей «Сцена» имеет приоритет 2 – то есть, желательно описать окружение персонажей, если после описания самих персонажей на это останется время. Более подробно уровни приоритетов представлены далее в таблице 6.

Таблица 6. «Описание приоритетов»

Приоритет	Важность	Сокращение
3	Необходимо описать	Сокращается при необходимости
2	Желательно описать	Сокращается при необходимости. Если сущности с приоритетом 3 уже сокращены, и сущность с приоритетом 2 сокращена, но они по-прежнему конфликтуют - описание исключается
1	Описать при возможности	Если сущности с приоритетами 2 и 3 уже сокращены, и сущность с приоритетом 1 сокращена, но они по-прежнему конфликтуют - описание исключается
0	Не описывается	Не описывается

Если после подбора синонимов и сокращений по приоритетам фразу до сих пор не удаётся уместить во временной интервал, система сокращает описание по своему усмотрению. В таком случае изначальная строка и итоговая строка (образованная системой) записывается в таблицу «Исключения» - для дальнейшего анализа администратором системы и её доработки.

Параллельная разработка алгоритма и базы данных системы позволила лучше понять разрабатываемую систему и связь данных внутри неё. В результате, помимо самого алгоритма и видов таблиц, была разработана схема базы данных (рисунок 5).

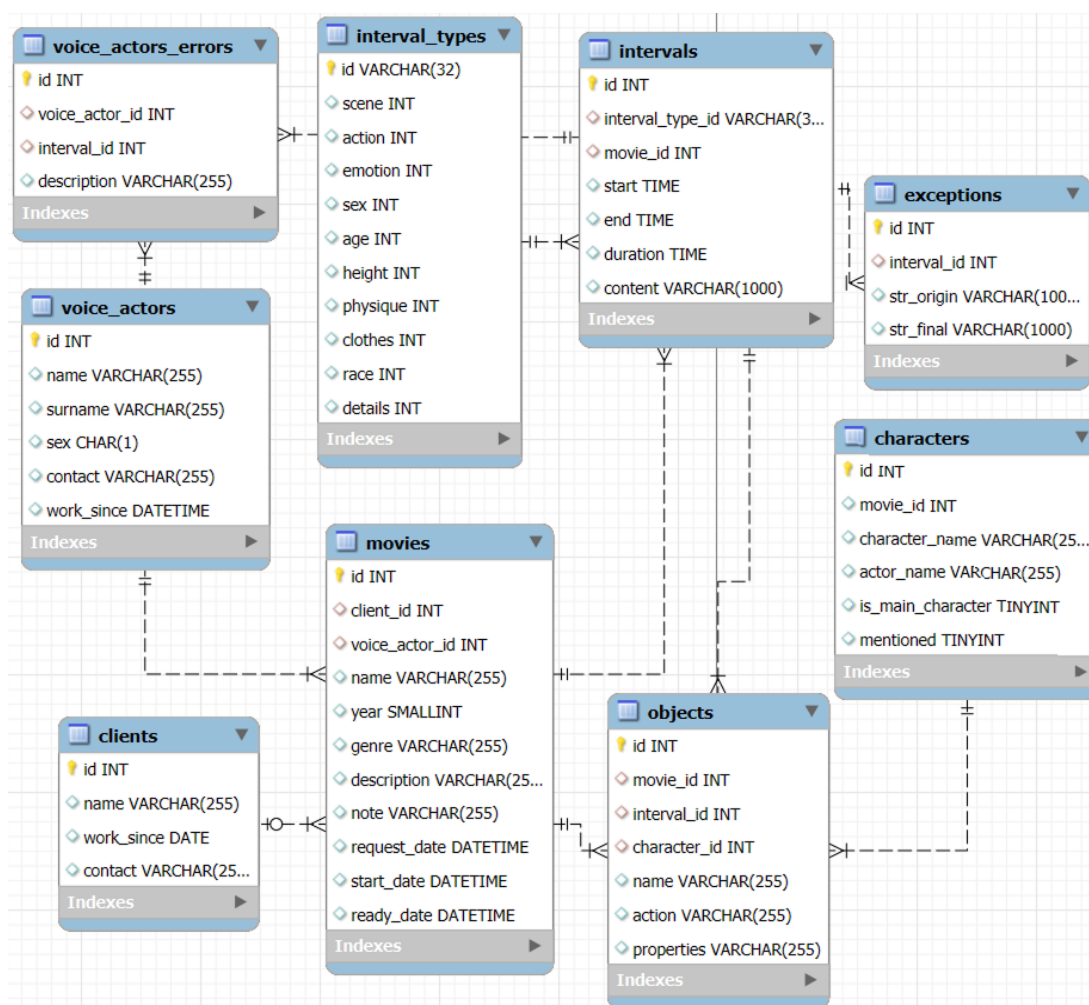


Рис. 5. Визуальная схема базы данных

4. Выводы

В рамках данного исследования была выполнена важная часть разработки информационной системы – разработка подробного алгоритма, учитывающего множество возможных сценариев работы системы.

Параллельная работа над алгоритмом и схемой базы данных позволила лучше представить работу реальной системы, отследить создание и преобразование информации от загрузки фильма в систему и до получения готовых текстовых описаний.

Собранная на прошлых этапах исследования теоретическая информация, построенный макет внешней (пользовательской) части системы, а также разработанная блок-схема алгоритма и схема базы данных позволяют заняться разработкой и интеграцией друг в друга обоих – внешней и внутренней – частей системы.

Литература

1. Тифлокомментаторы, комментирующие фильмы и спектакли для незрячих, – о своей работе – Афиша Daily. Взято из <https://daily.afisha.ru/context/24784-pomoch-nezryachim-lyudyam-posmotret-spektakl-ili-film-kto-takie-tiflokommentatory/>. Дата обращения: 20.05.2024.
2. Apostol Films. Кино с тифло. Взято из <https://apostolfilms.ru/кино-с-тифло/>. Дата обращения: 20.05.2024.
3. ГОСТ Р 59813-2021 «Тифлокомментирование». – Москва: Российский институт стандартизации, 2021. – 12 с.
4. Быстрый старт: Визуальное проектирование базы данных в MySQL Workbench. Взято из <https://habr.com/ru/articles/175985/>. Дата обращения: 23.05.2024
5. Правила построения блок-схемы алгоритма. Взято из <https://studfile.net/preview/1697967/page:42/>. Дата обращения: 25.05.2024.
6. Циклические алгоритмы, Энгельсский Технологический Институт. Взято из <http://techn.sstu.ru/kafedri/%D0%BF%D0%BE%D0%B4%D1%80%D0%B0%D0%B7%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D1%8F/1/MetMat/shaturn/prog/part7.htm>. Дата обращения: 26.05.2024.
7. Geekbrains: Блок-схема алгоритма: разбираемся в особенностях. Взято из <https://gb.ru/blog/blok-shema-algoritma/>. Дата обращения: 26.05.2024.
8. Тьма (сериал 2017-2020). Взято из https://www.kinopoisk.ru/series/1032606/?utm_referrer=www.kinopoisk.ru. Дата обращения: 26.05.2024