

Отчёт по теме: Мифические существа из игр и кино с поддержкой аудио и изображений

Выполнили:

Вязаницын Ф.А
Ярмолович М.А
Ильин Д.А

1. Предметная область

Исследование мифических существ в видеоиграх и кинематографе с акцентом на аудиовизуальные элементы:

- Анализ визуального дизайна существ.
- Изучение аудиосопровождения.

2. Задача для будущей системы

Тип системы: Мультимодальная вопросно-ответная система (ИИ), объединяющая текстовые, аудио- и визуальные данные.

Назначение:

1. Распознавание существ по изображениям или аудиозаписям (например, загруженный скриншот → информация о существе).
2. Предоставление списка похожих существ.

Зачем нужна эта система?

Решение практических проблем пользователей:

- Многие мифические существа в играх и кино имеют уникальный визуальный дизайн и звуковое сопровождение. Пользователи часто сталкиваются с ситуациями, когда:
 - Видят существо в игре или фильме, но не знают его названия или происхождения.
 - Слышат характерный звук (например, рык), но не могут его идентифицировать.
- Система решает эти проблемы, позволяя **распознавать существа по изображениям или аудио** и предоставляя о них информацию.

Углубление взаимодействия с контентом:

- Для фанатов и исследователей важно не только узнать название существа, но сравнить его с аналогами из других вселенных.

Поддержка творческих задач:

- Художники, гейм-дизайнеры могут использовать систему для поиска вдохновения. Например: найти похожих существ по стилю арта.

Почему для системы нужны данные?

- **Обучение мультимодального ИИ:**
 - Система должна **понимать связь между разными типами данных**:
 - **Изображения** → для обучения компьютерного зрения распознавать уникальные черты существ (например, форму крыльев дракона).
 - **Аудио** → для анализа звуковых особенностей (голос, темп, тональность).
- **Обеспечение точности и полнота**
- **Создание базы для сравнения:**
 - Для функции "похожие существа" необходима разметка данных по категориям:
 - **Визуальные признаки** (цвет, форма, стиль арта).
 - **Аудиопризнаки** (тональность, темп, жанровая принадлежность).

Источники данных:

<https://heartsfx.github.io/cards/all>

<https://freeanimalsounds.org/downloads/>

Аннотационная схема

Описание столбцов:

1. id

- Data Type: `int`
- Example Values: 1, 2, 150 (уникальный числовой идентификатор)

2. name

- Data Type: `str`
- Example Values: "Smaug the Dragon", "Hippogriff", "Demogorgon" (название существа / creature name)

3. universe

- Data Type: `str`
- Example Values: "The Hobbit", "Harry Potter", "Warcraft" (источник происхождения)

4. type

- Data Type: `list[str]`
- Example Values: ["dragon"], ["hybrid", "bird"], ["demon"] (тип существа)

5. sex

- Data Type: `str`
- Example Values: "male", "female", "hybrid", "unknown" (пол существа)

6. habitat

- Data Type: `list[str]`
- Example Values: ["mountains"], ["forest", "swamp"], ["dungeon"] (места обитания)

7. picture

- Data Type: `np.array[] / str`
- Example Values: "smaug.jpg", "https://example.com/demogorgon.png" (URL/путь к изображению)

8. body_parts

- Data Type: `list[str]`
- Example Values: ["wings", "tail"], ["tentacles", "fangs"] (особенности анатомии, можно добавить количество)

9. color

- Data Type: `list[str]`
- Example Values: `["red", "gold"], ["black"], ["transparent"]`
(доминирующие цвета)

10. size

- Data Type: `str`
- Example Values:
 - `"nano"`
 - `"small"`
 - `"medium"`
 - `"large"`
 - `"extra large"`
 - `"colossal"`

11. audio

- Data Type: `str`
- Example Values: `"roar.mp3", "https://example.com/scream.wav"`
(URL/путь к аудио)

12. audio_duration

- Data Type: `float`
- Example Values: `3.5, 12.0, 0.8` (длительность в секундах)

13. audio_tone

- Data Type: `str`
- Example Values: `"medium", "low", "high"` (тон звука)

1. Типы разметки (уровни разметки)

1. Семантическая разметка:

- `name, universe, type, habitat, sex`

2. Анатомическая/визуальная разметка:

- `body_parts, color, size, picture`

3. Акустическая разметка:

- `audio, audio_duration, audio_tone`

2. Модальность

- Текст: name, universe, type, habitat, sex, color, size
- Изображение: picture, визуальные признаки (цвет, анатомия)
- Аудио: audio, audio_duration, audio_tone
- **duration_sec** – длительность звука в секундах
- **rms_energy** – средняя энергия сигнала (громкость)
- **spectral_centroid_Hz** – спектральный центр (в Гц), отражающий "яркость" звука
- **spectral_bandwidth_Hz** – ширина спектра (в Гц), характеризующая диапазон частот

3. Визуализация

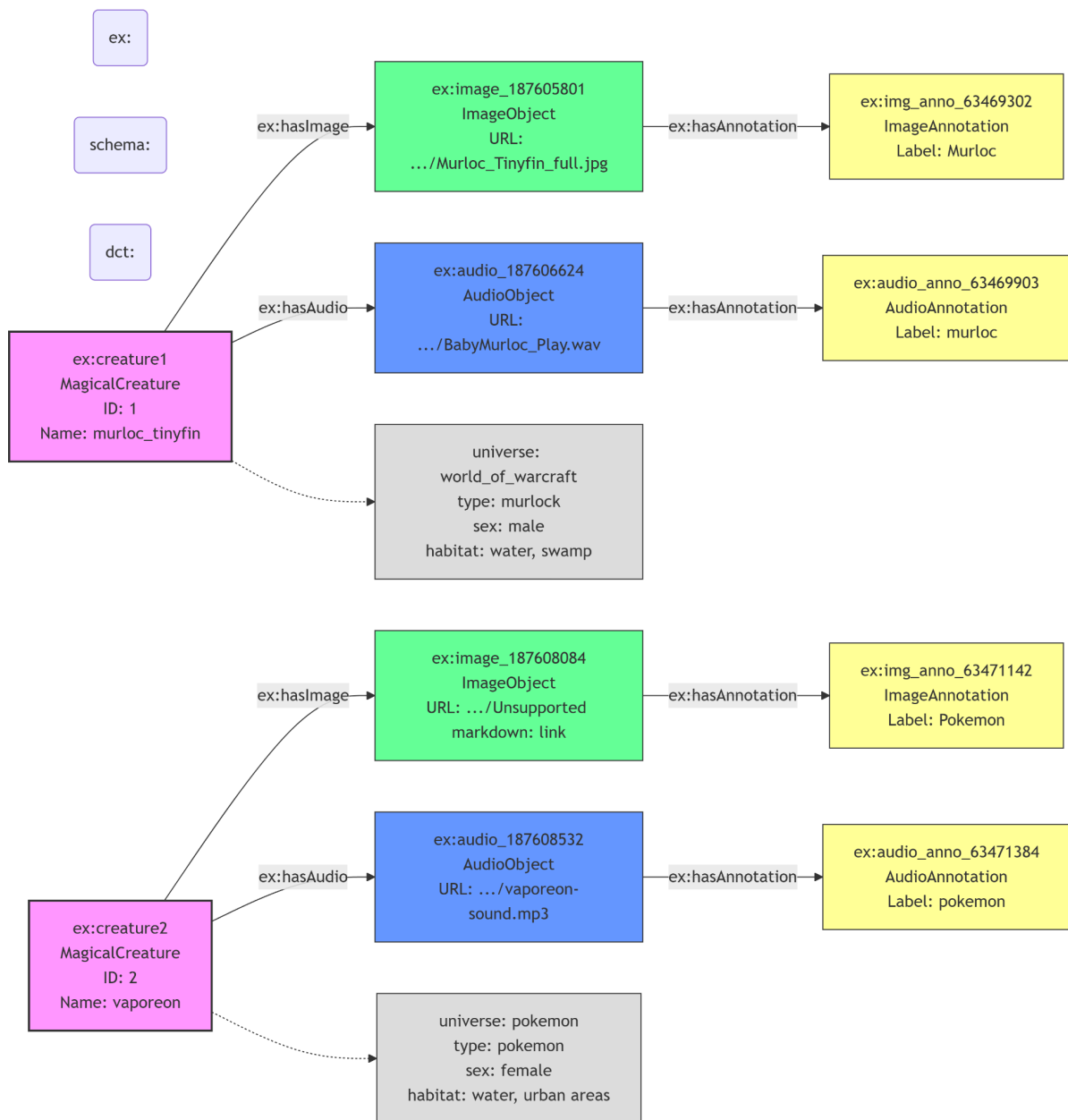
- Галерея изображений: миниатюры существ из поля picture
- Аудио-плеер: воспроизведение файлов audio с отображением audio_duration и описанием audio_tone
- Фильтрация и кластеризация:
 - По type, habitat, color, size
 - Группировка по universe (вселенной)
- Сравнение существ: визуальное и акустическое сопоставление (пример: драконы из разных вселенных)

4. Связующие сущности (для мультимодальной привязки)

- id — уникальный идентификатор для связывания всех модальностей
- name — основная сущность для человекочитаемой идентификации
- Кросс-модальные связи:

- `picture` ↔ `body_parts`, `color`, `size` (визуальная семантика)
- `audio` ↔ `audio_tone`, `audio_duration` (акустическая семантика)

Граф знаний:

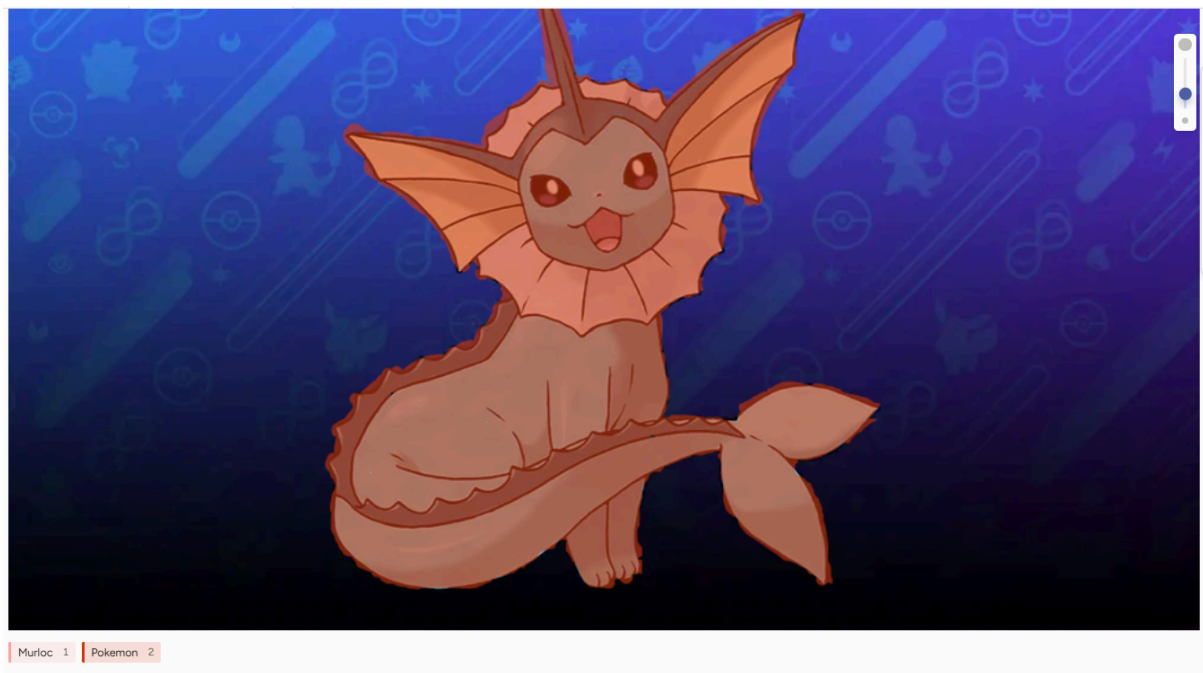


ссылка на ноутбук с графом знаний:

<https://colab.research.google.com/drive/13kZS0CTdRYO4WCEfrPbhPyjwnBvwHK6V?usp=sharing>

Уровень разметки	Тип разметки	Модальность	Формат разметки	Способ визуализации / Инструмент разметки	Связанные сущности
Классификация животного	тип мифического животного	Изображение	CSV / JSON	LabelImg / VGG Image Annotator / вручную	animal_id picture_id
Атрибуты внешности	Набор атрибутов (рога, крылья и пр.)	Изображение	CSV / JSON / XML	Label Studio / VGG Image Annotator	animal_id picture_id
Сегментация	Маска объекта на изображении	Изображение	PNG / COCO JSON	CVAT / LabelMe	animal_id picture_id
Тип звука	Категория звука (рев, шипение и пр.)	Аудио	CSV / JSON	Audacity / Sonic Visualiser / Label Studio	animal_id audio_id
Характеристики звука	Спектральные и временные параметры	Аудио	CSV / JSON	Audacity / Python (Librosa)	animal_id audio_id
Тематическая разметка текста	Темы (легенда, место обитания, способности)	Текстовые данные	JSON / CSV	Prodigy / Label Studio / вручную	animal_id

Разметка сущностей (NER)	Выделение сущностей (имена, места, эпохи)	Текстовые данные	CONLL / JSON / IOB	Prodigy / SpaCy / Label Studio	animal_id
--------------------------------	---	---------------------	--------------------------	--------------------------------------	-----------





Murloc 1 | Pokemon 2

