

Banco de Dados | Oblivion

Tudo relacionado ao banco de dados.

- [Modelo Entidade Relacionamento \(MER\)](#)
- [Modelo Lógico](#)
- [Modelo Físico](#)

Modelo Entidade Relacionamento (MER)

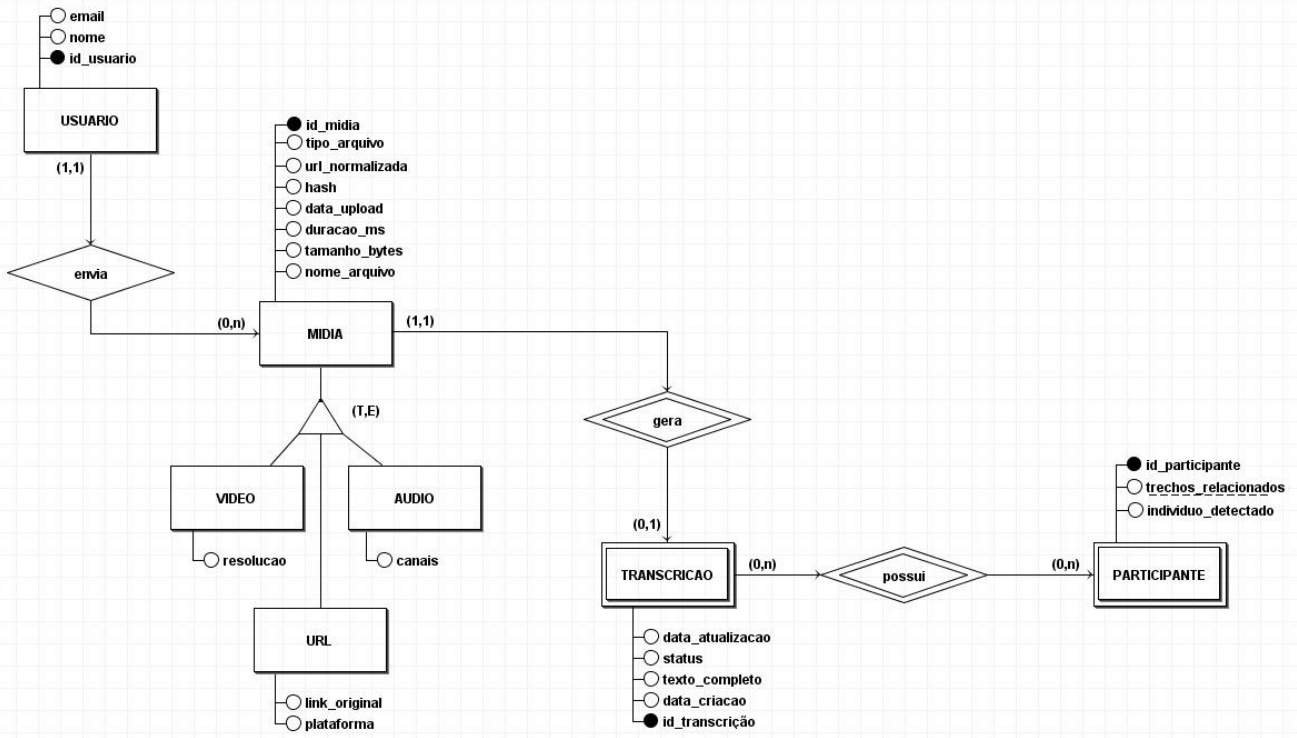
Sobre o Banco de Dados

Este é um modelo de banco de dados projetado para um **sistema de processamento e transcrição de mídias**.

O fluxo principal do sistema parece ser o seguinte:

1. Um **Usuário** envia (faz upload) de um arquivo de **Mídia**.
2. Essa **Mídia** pode ser de três tipos específicos: um **Vídeo**, um **Áudio** ou um link de uma **URL** externa (como um vídeo do YouTube).
3. O sistema processa a **Mídia** e **gera** uma **Transcrição** do seu conteúdo.
4. Durante a transcrição, o sistema identifica os diferentes **Participantes** (vozes ou pessoas) na mídia e os associa à transcrição correspondente.

É um sistema robusto para serviços que convertem fala em texto (Speech-to-Text), legendagem automática, ou análise de conteúdo de áudio e vídeo.



Modelo Lógico

Modelo Lógico (MongoDB) - Projeto OBLIVION

Este documento descreve a estrutura das coleções e documentos para o banco de dados MongoDB do projeto OBLIVION. Ele serve como a "fonte da verdade" para a estrutura de dados que será implementada.

Coleção: `usuarios`

Armazena os dados de cada usuário cadastrado no sistema.

Campo	Tipo	Descrição	Observações
<code>_id</code>	ObjectId	Identificador único gerado pelo MongoDB.	Chave Primária.
<code>nome</code>	String	Nome completo do usuário.	Obrigatório.
<code>email</code>	String	Email do usuário para login e contato.	Obrigatório, Índice Único .
<code>dataCadastro</code>	ISODate	Data e hora do cadastro do usuário.	Obrigatório.

Coleção: `midias`

Armazena os metadados de qualquer fonte de mídia enviada (vídeo, áudio ou URL), implementando a generalização através do campo `tipo`.

Campo	Tipo	Descrição	Observações
<code>_id</code>	ObjectId	Identificador único da mídia.	Chave Primária.

Campo	Tipo	Descrição	Observações
<code>usuarioid</code>	ObjectId	Referência ao <code>_id</code> do usuário que fez o upload.	Chave Estrangeira (Obrigatório).
<code>hash</code>	String	Hash (ex: SHA-256) do conteúdo do arquivo.	Índice Único. Nulo se <code>tipo</code> = "URL".
<code>urlNormalizada</code>	String	URL padronizada para evitar duplicidade de links.	Índice Único. Existe apenas se <code>tipo</code> = "URL".
<code>tipo</code>	String	Tipo da mídia: "VIDEO", "AUDIO", ou "URL".	Obrigatório.
<code>nomeArquivo</code>	String	Nome original do arquivo enviado.	Não aplicável se <code>tipo</code> = "URL".
<code>tamanhoBytes</code>	Number	Tamanho do arquivo em bytes.	
<code>tipoArquivo</code>	String	MIME type do arquivo (ex: "video/mp4").	Não aplicável se <code>tipo</code> = "URL".
<code>duracaoMs</code>	Number	Duração do conteúdo em milissegundos.	
<code>dataUpload</code>	ISODate	Data e hora do envio.	Obrigatório.
<code>resolucao</code>	String	Resolução do vídeo (ex: "1920x1080").	Opcional (existe apenas se <code>tipo</code> = "VIDEO").
<code>canais</code>	String	Canais do áudio (ex: "stereo").	Opcional (existe apenas se <code>tipo</code> = "AUDIO").
<code>linkOriginal</code>	String	A URL original enviada pelo usuário.	Opcional (existe apenas se <code>tipo</code> = "URL").
<code>plataforma</code>	String	Plataforma de origem da URL (ex: "YouTube").	Opcional (existe apenas se <code>tipo</code> = "URL").

Coleção: `transcricoes`

Armazena o estado e o resultado de cada processo de transcrição.

Campo	Tipo	Descrição	Observações
<code>_id</code>	ObjectId	Identificador único da transcrição.	Chave Primária.
<code>midiaId</code>	ObjectId	Referência à <code>_id</code> da mídia que está sendo transcrita.	Chave Estrangeira (Obrigatório). Índice.
<code>status</code>	String	Status do job: "pendente", "processando", "concluido", "erro".	Obrigatório. Índice.

Campo	Tipo	Descrição	Observações
textoCompleto	String	O texto final da transcrição.	
dataCriacao	ISODate	Data e hora do início do job.	Obrigatório.
dataAtualizacao	ISODate	Data e hora da última atualização de status.	Obrigatório.
participantes	Array de Objetos	Lista dos participantes identificados no texto.	Ver schema embutido abaixo.

Schema Embutido: participantes

O campo `participantes` é um array de objetos, onde cada objeto segue a estrutura:

Campo	Tipo	Descrição
individuoDetectado	String	Nome/rótulo do locutor (ex: "Locutor A").
trechosRelacionados	Array de Objetos	Lista de trechos de fala daquele indivíduo.

Modelo Físico