

Ferramentas de análise de dados em bases ISIS

Gustavo Gannam - Cinemateca Brasileira



Figure 1:

Novembro de 2017

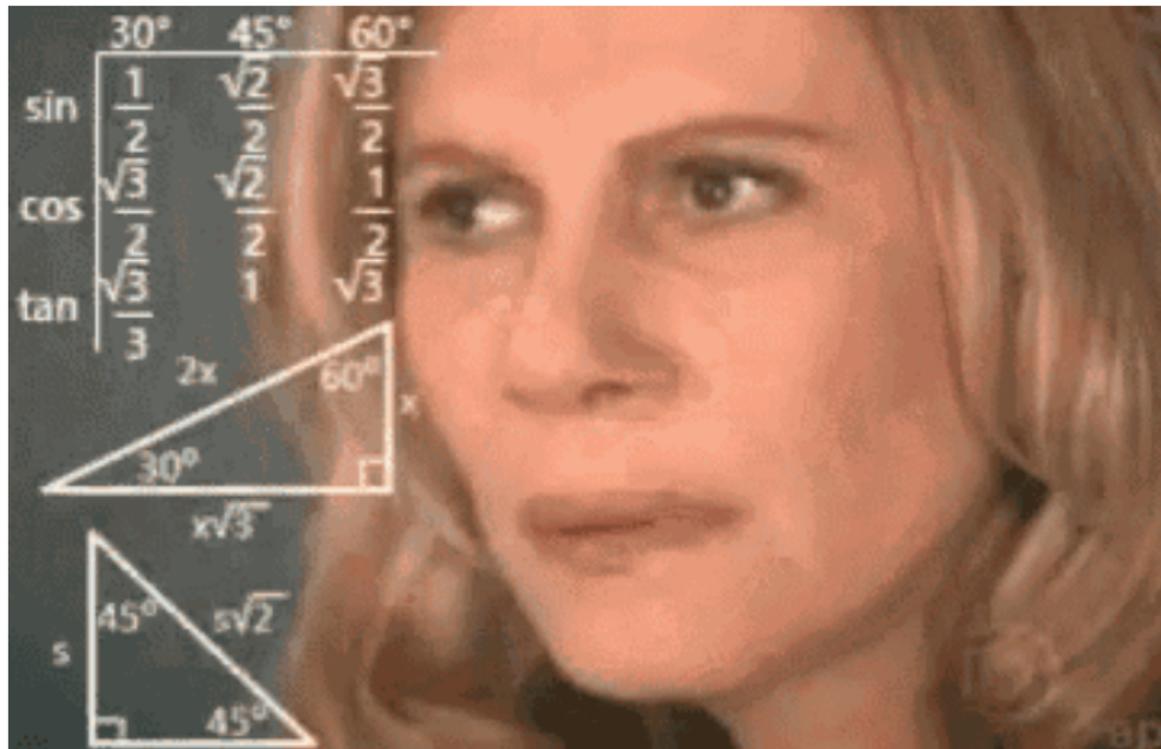
Diagnóstico comum

Vários bancos de dados e diversas soluções que não estão integrados.
A Cinemateca utiliza vários tipos de banco de dados:

- ▶ ISIS: catalogação e movimentação do acervo
- ▶ SQLite: documentação interna
- ▶ MySQL: sites como bcc.gov.br e cinemateca.gov.br
- ▶ CSV: planilhas

Sem integração de dados. . .

“Levando em consideração os diversos materiais de uma obra audiovisual, qual é o seu estado técnico de preservação?”



Objetivo da pesquisa

Desenvolver ou integrar uma ferramenta com uma interface moderna, para que seja facilitada a pesquisa pelos diferentes núcleos da instituição nos bancos de dados da **Cinemateca Brasileira**.
Requisitos mínimos: a solução em **software livre**, contemplando os dados ISIS, sem necessidade prévia de migração.

Banco de dados MongoDB

Utilizando ferramentas em **software livre**, transformamos a base ISIS em um arquivo **formato JSON** e importamos no **MongoDB**. O **MongoDB** é um banco de dados não relacional, isto é, **semi-estruturado**, baseado em **coleções** e **documentos** e não em tabelas e linhas (SQL).

Ferramentas de análise de dados

- ▶ **Apache Drill** - drill.apache.com
- ▶ **Redash** - redash.io
- ▶ **Presto** - prestodb.io
- ▶ **Metabase** - metabase.com

Metabase

- ▶ Possui um construtor de pesquisa (*query builder*) muito amigável para o usuário
- ▶ Visualização dos registros numa tabela que pode ser exportada (XLSX, CSV, JSON)
- ▶ É possível salvar as pesquisas (*queries*), assim como fazer coleções
- ▶ É rápido criar gráficos e visualizar os dados de outras formas
- ▶ Os resultados podem ser salvos e exibidos num painel (público ou privado)

Limitações

- ▶ O suporte ao MongoDB não é o seu foco principal. . .
- ▶ Ainda não é possível fazer cruzamento (*subqueries*) entre as bases do MongoDB
- ▶ Há casos em que não é possível pesquisar dentro de determinados subcampos (JSON com campos vazios e/ou nulos)
- ▶ Embora muito estável, o Metabase ainda está nos estágios iniciais de desenvolvimento (versão 0.26.x)

Resultados

- ▶ Facilita a visualização de conjuntos de informações das bases ISIS
- ▶ Melhor fluxo de trabalho. **Exemplo:** Geração de relatórios de status de preservação dos materiais fílmicos
- ▶ Resultados de pesquisas complexas em menor tempo
- ▶ Aumento na qualidade da análise das informação (inconsistências, erros de preenchimento, falta de padronização)

Documentação

Em breve no nosso site e git.

Outros projetos

- ▶ **isislog:** versionamento e registro de mudanças em bases ISIS.
- ▶ **climapi:** monitoramento online da climatização do seu acervo com Raspberry Pi.

Contato: **dev@cinemateca.gov.br**

<http://cinemateca.gov.br> | <http://bcc.gov.br>