Skill Share

Simon Wörpel & Alex Ștefănescu

25.05.2025

Who are we?

Simon Wörpel

- Director of Technology, DARC

Alex Ștefănescu

- Senior Developer, DARC

How to reveal metadata and other hidden information in PDF documents

Should I open this PDF?

Two options:

- OpenAleph
  - also enables search, cross-referencing
- dangerzone.rocks
  - transforms PDF into a series of images (no code, no surprises when clicking links)

Metadata

Command line: exiftool

- Author, e-mail, timestamps, software
- **exiftool file.pdf**

Remove Metadata

Metadata can be edited with **exiftool**, but **all edits are reversible**.

This includes removing metadata.

exiftool -all= some.pdf
qpdf --linearize some.pdf done.pdf

Experiment: remove redactions

qpdf --qdf --object-streams=disable initial.pdf result.pdf

The PDF should now be “readable” in a text editor.

Attempt to remove the black “redaction” boxes by deleting lines that look like coordinates for a drawn rectangle. Consult the resulting PDF file in your PDF viewer app.

AI-powered search through audio and video

Why tell when you can show?

Our goal will be to:

- Upload audio files to OpenAleph
- Find all files that mention a name
- Read the transcripts of the files

Why tell when you can show?

1. Upload audio to OpenAleph
2. Transcribe it locally
3. Get FTM entities from the Postgres database
4. Get the FTM entities from OpenAleph
5. Add the transcription to the entities
6. Push the new entities to OpenAleph

Join the conversation

darc.social

- Join the conversation
- Suggest new features
- How are you using OpenAleph in your workflow?

Stay up to date

openaleph.org/blog

- New features
- Technical behind-the-scenes

darc.li/news

- Newsletter

Thanks!

https://darc.li/dh25-skillshare