

JHOVE

Der Evergreen der LZA-Hitparade?

Yvonne Tunnat

nestor Workshop Formaterkennung

29.06.2017



Was ist JHOVE?

- Erstmals veröffentlicht 2005 von der Harvard University Library
- Seit März 2014 gehostet bei der OPF (Open Preservation Foundation)
- Community hilft mit via GitHub, Document Interest Group, JHOVE Hack days
- JHOVE kann 14 verschiedene Formate validieren, u. a. PDF, TIFF, JPEG

Drei verschiedene Nutzungsmöglichkeiten:

- GUI (Graphical User Interface)
- via Kommandozeile (somit als Batch-Programm)
- Java library

Kommandozeile

Eingabeaufforderung öffnen im Ordner, in dem JHOVE sich befindet

```
jhove -h (zeigt alle Module und Optionen an)
```

```
jhove datei.pdf -o (gibt Ergebnis aus)
```

```
jhove datei.pdf -o > ergebnis.txt (gibt Ergebnis in  
Textdatei aus)
```

Da die Textdatei locker über 800 Zeilen haben kann, ist das aber nicht sehr übersichtlich (*Beispiel siehe Stick Portable -> JHOVE -> „ergebnis_jhove.txt“*)

JHOVE (Version 1.16.3)

Software-Repository -> JHOVE -> jhove-gui.bat

Ggf. über Edit -> select Module das entsprechende Modul auswählen, z. B. PDF-hul. Andernfalls besteht auch die Gefahr, dass ein invalides PDF einfach als „well-formed and valid Bytestream“ ausgezeichnet wird.

Via „File“ -> open oder einfach mit Drag und Drop eine Datei oder einen Ordner zum Analysieren auswählen

Output: auch speicherbar und mit Editor zu öffnen

JHOVE als Library nutzen

Java-Entwicklerinnen oder Entwickler im Raum?

JHOVE Tips for developers:

<https://www.smashwords.com/books/view/270759>

Da ist sehr gut erklärt wie man die JHOVE java library nutzen kann.

Hands On

- Zeit zum Ausprobieren, Testen und Fragen



Diskussion JHOVE

- False positives und false negatives
- Output
- JHOVE integrieren

Weiterführende Links

GitHub JHOVE: <https://github.com/openpreserve/jhove>

OPF JHOVE: <http://openpreservation.org/technology/products/jhove/>