

Metadata Tool

In dem Modul Management of Scientific Data habe ich dieses Semester zusammen mit einer Gruppe Studenten und einem Team von Geologen einen Data Management Plan ausgearbeitet für das Projekt SALDi (Südafrika Landdegradationsmonitor).

Dabei geht es um die langfristige Untersuchung von Bodenerosion in Südafrika. Das Forscherteam sammelt dafür verschiedenste Daten von mehreren Orten (von Satellitendaten, Feuchtigkeitsdaten und historischen Luftbildern bis hin zu Sonardaten).

Ich hab mich dabei besonders mit den entstehenden Metadaten auseinandergesetzt. Da es sich um solch verschiedene Dateiformate und Daten handelt kommen dementsprechend spezielle Metadaten vor. Die Gruppe hatte jedoch bis jetzt wenig allgemeine Meta- und Prozessdaten gesammelt. Der von uns vorgeschlagene Standard sieht vor zu jeder Datei in einem extra Verzeichnis eine Metadaten Datei anzulegen:

```
sites/
├── site1/
│   ├── test.css
│   ├── test.txt
│   └── metadata/
│       ├── test-css-metadata.txt
│       └── test-txt-metadata.txt
└── site2/
    ├── test2.txt
    ├── metadata/
    │   └── test2-txt-metadata.txt
    ├── fonts/
    │   ├── test2.csv
    │   └── metadata/
    │       └── test2-csv-metadata.txt
```

In diesen Dateien soll jeweils eine Liste von metadata tags mit zugehörigen Inhalt stehen. Ich würde gerne als Python Abschlussprojekt ein Tool schreiben womit man Metadaten in diesem Standard verwalten und editieren kann.

Was das Tool leisten soll

- Beim Start soll der Datenordner ausgewählt werden (open file dialog)
- Daraus soll eine liste aller Daten erzeugt werden die in diesem Ordner liegen
 - rekursiv
- diese Liste soll in einer grafischen Oberfläche dargestellt werden
- klickt man ein Element der Liste an, sollen die Metadaten als Textfelder angezeigt werden
 - Wenn schon welche vorhanden sind sollen diese dargestellt werden
 - werden die Metadaten verändert sollen sie nach einer Bestätigung in die Metadaten Dateien geschrieben werden

- es soll Möglich sein die Liste der Keywords zu bearbeiten
 - beinhaltet managen der Liste als Textdokument im Tool Verzeichnis
- es soll Möglich sein Einträge anderer Dateien zu übernehmen
 - ggf. für den gesamten Unterordner
- Ein besonderer Punkt nimmt die Beschreibung der Bearbeitungsprozesse/Erzeugungsprozesse die zu diesen Daten geführt haben, darin soll der Ablauf der Verarbeitungsschritte aufgelistet werden
 - es sollen neue solcher Bearbeitungsprozesse definiert werden können, die dann später aus einer Liste für die jeweilige Datei ausgewählt werden

Besondere Schwierigkeiten die ich sehe

- das Darstellen der einzelnen Textboxen für die Metadata keywords, da es davon im Prinzip beliebig viele geben können soll
 - eventuell mehrere Seiten, Tabs oder eine scrollbare Liste
- das Speichern der Verarbeitungsprozesse sowie der Keywords
 - Jeder hat nur seine lokalen Änderungen
- Die Darstellung der Liste der Dateien mit den jeweiligen Unter und Oberverzeichnissen
 - eventuell wie ein dateibrowser

Generelle Idee des Aufbaus

- eine Klasse um das Bearbeiten der Dateien zu bewerkstelligen
 - das restliche Programm sollte dann mit den daraus erzeugten Objekten umgehen können wie mit dictionaries
- eine Datenstruktur um alle Dateien-Objekte in dem Pfad zu verwalten
- Globale Liste von Tags und Bearbeitungsprozessen
 - werden überschrieben falls neue hinzugefügt werden
- GUI in tkinter → wird erst später genauer spezifiziert wenn das Gerüst steht