

nestor-AG Formaterkennung, 7. Treffen

14. März 2017 im Landesarchiv Nordrhein-Westfalen, Duisburg, 11:00 bis 16:00 Uhr

Anwesend: Andre, Heinz, Konrad, Michelle, Martin (Protokoll), Yvonne.

Entschuldigt: Christine, Claire, Marion, Mario, Stefan, Svenia, Tim.

0 Begrüßung

1 Aktueller Stand aller Arbeitspakete

AP 1 Tool Registries (Yvonne in Vertretung für Claire)

Sofern neue Erfahrungen mit Tools dazukommen, tragen alle Mitglieder der AG diese stets in die Tabelle im internen Wiki ein.

Die Zuarbeit zu COPTR ist weiterhin Thema für alle Mitglieder der AG. Michelle hat eine Liste mit Kandidaten für fehlende Beiträge, die im Wiki bereitgestellt wird.

Christine ist in Elternzeit bis voraussichtlich April/Mai 2018. Die Vorbereitung der Umfrage wird solange vertagt.

AP 2 Testsuites (Andre/Yvonne)

Yvonne hat auf GitHub eine Organisation namens "nestorFormatGroup" erstellt (<https://github.com/nestorFormatGroup>), weil gemeinsames Arbeiten mit sowie externes Referenzieren von Source Code, Testdateien u.ä. so einfacher ist und eine auch international größere Reichweite vermutet wird als im nestor-Wiki. Jedes AG-Mitglied (mit einem GitHub-Account) kann mitwirken. Künftig sollen geeignete Daten bevorzugt auf GitHub gesammelt werden.

Mittlerweile gibt es dort ein Abbild der bisher gesammelten Testdateien. Vorerst hat Yvonne die Dateien in fünf zip-Ordner (MS Access, MS Word, MS Excel, Textdateien, Office Files (für alles andere)) aufgeteilt. Sie wird das Repository aber noch differenzierter strukturieren.

Außerdem hat Yvonne aus den OPF Govdocs und unserer Testsuite sowie Github eine Testsuite mit genau 100 möglichst verschiedenen Dateien zusammengestellt, die vielleicht für den Formaterkennungs-Workshop beim nestor-Praktikertag nützlich sind.

AP 3 Register und Kontaktbörse für obsoleszente Speichertechniken

(»R.O.S.T.«) (Heinz)

Heinz hat die geplante Umfrage in LimeSurvey umgesetzt, was sehr gut aussieht. In der gehosteten Variante auf limequery.org sind 25 Antworten pro Monat gratis, realistischer wäre aber ein Abo für 30 Euro pro Monat mit zunächst etwa drei Monaten Laufzeit für die Dauer der Umfrage. Yvonne klärt mit der Geschäftsstelle, ob nestor diese Kosten übernehmen kann. Parallel prüft Andre den Aufwand, LimeSurvey kostenfrei auf Gesis-Servern zu hosten (mit der Option auf eine schicke Domain-Weiterleitung unter der langzeitarchivierung.de-Flagge).

Michelle hat Kontakt zu dem Projekt "Mediapeda: Physical Format Carrier Resource" (National Library of Australia) aufgenommen und nach der Möglichkeit für Links auf Holographen und systematische Suchoptionen gefragt sowie Mitarbeit der AG angeboten. Die Antwort war freundlich und interessiert, aber hinsichtlich der gestellten Fragen noch nicht detailliert genug. Heinz fragt noch einmal nach.

AP 4 Format Registries (Martin)

Die Anleitung für die Zuarbeit zu PRONOM ist i.W. fertig. Martin hat das bisherige Feedback eingearbeitet und wird noch einen kurzen Abschnitt zur Beschreibung von Bitmustern durch reguläre Ausdrücke ergänzen. Anschließend kann die Anleitung ins öffentliche Wiki verschoben und beworben werden.

AP 5 Formatvalidierung (Yvonne in Vertretung für Stefan)

Es ist inzwischen eine ausreichende Menge von PDF/A-Dateien für das geplante Software-Benchmarking der Validatoren vorhanden. Claire füttert die verschiedenen Validierungstools damit und leitet die Ergebnisse zur Auswertung an die AG weiter. Das weitere Vorgehen wird anschließend abgestimmt.

AP 6 Best Practices in der Identifizierung (Yvonne in Vertretung für Claire)

Yvonne entwirft einen Kurzaufsatz DROID vs. Siegfried für die nestor-Reihe. Michelle und Martin liefern Feedback.

2 Vorträge beim nestor-Praktikertag

Der diesjährige Praktikertag zum Thema "Formaterkennung, Formatvalidierung und Tools" findet am Mittwoch, 28.06.2017 von 09:30 bis 16:30 Uhr in der ZBW in Kiel statt.

Neben Begrüßungen, Austauschrunden und Kurzvorträgen der Aussteller wird das inhaltliche Programm von Ange Albertini ("on file formats"), Claire Roethlisberger ("Sinn und Zweck der Formaterkennung und deren Grenzen"), Marco Klindt ("Bist Du Dir sicher? JPEG/JFIF, JPEG/Exif und TIFF/Baseline überprüfen"), Michelle Lindlar ("Wahrheit oder Pflicht? Validierung mit JHOVE") und Dietrich von Seggern ("PDF und PDF/A Validierung - im Brennpunkt Schriften") gestaltet.

Details zur Planung stehen im internen Wiki auf der Seite zum Praktikertag 2017.

3 Vorbereitung des Workshops nach dem nestor-Praktikertag

Quasi als Fortsetzung des Praktikertags veranstaltet die AG am folgenden Tag, also am Donnerstag, 29.06.2017 von 10:00 bis 15:00 Uhr, ebenfalls in der ZBW in Kiel, einen praktisch orientierten Workshop zum selben Thema.

Es werden nach einem Überblick zur Funktionsweise von Formaterkennungs- und Formatvalidierungstools in kurzer Taktung verschiedene solcher Tools praktisch vorgestellt. Diese Beiträge sollen eine relativ einheitliche Struktur haben, damit ein Vergleich der Tools leichter fällt (was tut es und was nicht, wer entwickelt es, was steckt man rein, was kommt raus, wie wird es bedient, usw.). Die Tools sowie Testdateien werden den Teilnehmenden auf USB-Sticks (beschafft die ZBW) kostenlos zur Verfügung gestellt. Auf zeitliche Blöcke für eigene Versuche mit den Tools

während des Workshops wird aus Zeitgründen verzichtet, damit nicht eine defekte Java-Installation auf einem einzelnen Laptop alle anderen aufhält.

Details zur Planung stehen im internen Wiki auf der Seite zum Praktikertag 2017.

4 Beitrag zum nestor-Leitfaden/Handbuch

Yvonne hat mittlerweile den Leitfaden durchgelesen und findet das Thema Formaterkennung da nicht passend. Daher hat sie mit Herrn Keitel gesprochen und eine Beteiligung "abgesagt", was dieser für die richtige Entscheidung hält. Stattdessen soll nach dem Praktikertag eine Aktualisierung der entsprechenden Anteile des nestor-Handbuchs erwogen werden.

5 Sonstiges

Das nächste AG-Treffen findet zur Vorbereitung von Praktikertag und Workshop am Dienstag, 27.06.2017, 14:30-18:30 Uhr in der ZBW in Kiel statt.

Für eine Veranstaltung der Reihe "nestor für newbies" am 16.05.2017 in Frankfurt wird jemand gesucht, der zur Dateivalidierung für Einsteiger spricht. In der AG findet sich leider niemand mit Zeit und ausreichender Kenntnis. Michelle fragt stattdessen Franziska Schwab (TIB).