

Bericht Tech Trifecta Week 2012

Oliver Goldschmidt

Im Rahmen der Förderung von BI International konnte ich vom 15.-18.10.2012 an der Tech Trifecta Week in Villanova/Pennsylvania/USA teilnehmen. Die Tech Trifecta Week bestand aus drei Veranstaltungen: der VuFind Summit, das Code4Lib Mid Atlantic Regionaltreffen und VuStuff. Ich besuchte alle drei Veranstaltungen.

VuFind Summit, 15. - 16.10.2012

Der VuFind-Summit fand 2012 zum zweiten mal statt. Die erste VuFind-Konferenz fand im September/Oktober 2010 unter dem Namen VuFind 2.0 Conference statt.

Der diesjährige Summit war geprägt von der Vorstellung einzelner VuFind-Lösungen, Vorträgen zur neuen VuFind-Version 2.0 sowie Überlegungen, einen zentralen Index aufzubauen. Alle Vorträge sowie die Folien für einige der Vorträge sind unter http://vufind.org/wiki/vufind_summit_2012 frei verfügbar.

Zentraler Index

Joe Lucia, Direktor der Falvey Memorial Library an der Villanova University erklärt in seiner Eingangsrede, dass VuFind ein erfolgreiches Produkt ist, das weltweit eingesetzt und weiterentwickelt wird. Er bezeichnet es als eine „virtual neighborhood“; alle VuFind-Anwender, auch wenn sie in gänzlich verschiedenen Gegenden der Erde sitzen, nutzen das gleiche Produkt und sind damit virtuell Nachbarn.

Er plädiert für den Einsatz von OpenSource-Software, weil es seiner Ansicht nach andere Wege gibt, Geld zu verdienen, als den Code, den man produziert, zu verschließen.

Die Villanova University nutzt Summon als zusätzlichen Index, der teuer bezahlt werden muss. Es sei an der Zeit, die Dinge selbst in die Hand zu nehmen und nicht mehr externe Indizes einzubinden, die bezahlt werden müssen, sondern eine nicht-kommerzielle Eigenlösung zu nutzen. Er fordert eine Umverteilung des Geldes; von dem Geld, das jährlich an externe Indexbetreiber fließt, könnte eine schöne Eigenentwicklung finanziert werden.

Beim Brainstorming zum Thema Central Index wird auf verschiedene bereits bestehende Produkte hingewiesen:

- LodLan
- GoKb (Global Open Knowledge Base)
- LibraryCloud

Ziel sollte es sein, einen Partner zu finden, der dabei helfen kann, die vorhandenen Daten zu bündeln und bereitzustellen.

Fraglich sind dabei die genauen Inhalte. Denkbar wären zum Beispiel Umlaufdaten bzw. soziale Empfehlungsdaten, oder auch ein bibliographischer Index. Der bibliographische Index findet auf Antriebe etwas mehr Interessenten. Die Form ist dabei unklar: neben der traditionellen indexbasierten Struktur (Solr) käme auch eine LOD (Linked Open Data)-basierte Lösung infrage. Die Organisation eines solchen Projektes wird sich jedoch als etwas schwierig erweisen: Lizenzdinge müssen berücksichtigt werden, es muss Verhandlungsführer gegenüber den Verlagen geben,...

Google Scholar scheidet als Datenlieferant aus, weil die Schnittstellen nicht offen und zuverlässig verfügbar sind.

Demian Katz, Chefentwickler des VuFind-Projektes an der Villanova University, ruft in diesem Zusammenhang alle Anwesenden und Interessierten auf, sich näher mit Linked Open Data zu beschäftigen. Diese Technologien werden seiner Ansicht nach in Zukunft noch wichtiger werden.

Eine wichtige Frage, die sich bei den Überlegungen um einen zentralen Index stellt, ist die Einbindung des Index in die VuFind-Oberfläche. Zu diesem Thema gab es mehrere Vorträge. An der TUBHH verfolgen wir einen Ansatz, alle Ergebnisse verschiedener Indizes in einer gemeinsamen Ergebnisliste darzustellen. Den gleichen Ansatz fahren auch die Kollegen im finc-Projekt an der UB Leipzig. Vor Kurzem gab es eine interessante Diskussion zu dem Thema in Anne Christensens Blog *A Growing Organism* (<http://xenzen.wordpress.com/2012/10/04/to-tab-or-not-to-tab-eine-discovery-gretchenfrage/>). Andere Hochschulen bevorzugen getrennte Ergebnislisten.

So berichten zum Beispiel **Chelsey Lobdell** und **Spencer Lamm** vom Swarthmore College/Tri-College-Consortium in den USA über die Einbindung von Primo Central über die API in VuFind. Nutzerstudien vor Ort haben ergeben, dass die „interfiled result list“ nicht gut ankam bei den Nutzern. Daher wurde der Weg beschritten, die Ergebnislisten in verschiedenen Tabulatoren darzustellen, so wie die Villanova University es auch tut. Es wurde ein neuer Recordset Handler geschrieben, der die API-Rückgabe (XML) parst und das Ergebnis an VuFind durchreicht.

An der Universitätsbibliothek Leipzig, wo ebenfalls VuFind mit Primo Central gekoppelt wird, wird die Primo Central Bridge benutzt. **Leander Seige** vom finc-Projekt an der Universitätsbibliothek Leipzig stellt die dortige Lösung mit der PC Bridge näher vor. Er geht in seinem Vortrag auch auf die Behandlung von Normdaten ein, die seiner Ansicht nach problematisch ist, weil für den Nutzer intransparent die Suchanfrage manipuliert wird.

Die Frage „Eine oder mehrere Ergebnislisten?“ wird auch in der Breakout-Session „Usability Testing“ auf der Konferenz thematisiert. Es wird die Frage aufgeworfen, wie man die verschiedenen Tabulatoren sinnvoll benennen soll. Zudem bestätigt **Björn Muschall** von der Universitätsbibliothek Leipzig, dass sie bei ihren Userbefragungen (anders als Chelsea) herausgefunden haben, dass die User eine kombinierte Ergebnisliste bevorzugen.

In meinem eigenen Vortrag (**Oliver Goldschmidt** von der Universitätsbibliothek der Technischen Universität Hamburg-Harburg) stelle ich Möglichkeiten vor, Indizes mit der Solr-Technik des Sharding miteinander zu kombinieren und dadurch mehrere Indizes in einer VuFind-Instanz oder gar in einer gemeinsamen Ergebnisliste darzustellen. Es gibt dazu zwei verschiedene Ansätze: mit dem ersten (der auch an der TUBHH verwendet wird) wird eine gemeinsame Ergebnisliste produziert, indem die einzelnen Indizes auf Solr-Ebene (Suchmaschinenebene) miteinander gekoppelt werden. Dazu ist es notwendig, dass die Indizes gleichartig strukturiert sein müssen (kleine Unterschiede können ausgeglichen werden, das birgt jedoch Risiken). Sollen verschiedene Ergebnislisten produziert werden, so können verschiedene Indizes in VuFind einkonfiguriert werden, die dann wahlweise von verschiedenen Ergebnis-Registerkarten genutzt werden. Auf diese Weise arbeitet das Bibliotheks-Servicezentrum (BSZ) in Baden-Württemberg mit verschiedenen Shards.

Die Sharding-Funktionen von VuFind, die von uns in Hamburg ursprünglich implementiert wurden, könnten als technische Grundlage genutzt werden, wenn es daran geht, einen externen Index zum lokalen Index zuzuschalten.

VuFind 2.0

Demian Katz stellt die Änderungen von VuFind 2.0 vor. Ein Ziel von VuFind 2.0 ist eine Codebereinigung. VuFind 2.0 enthält einen Plugin-Layer, der eine verbesserte Erweiterbarkeit ermöglicht. Zudem basiert VuFind 2.0 auf der neuen PHP-Version und dem Zend Framework 2, so dass der veraltete Kern der Software auf modernen Technologien aufbaut.

Er stellt anschließend die Roadmap dar und berichtet, inwieweit die Punkte, die auf dem letzten Summit im Jahr 2010 beschlossen wurden, realisiert wurden. Dabei stellt er fest, dass die Punkte

nicht systematisch abgearbeitet wurden, sondern jetzt in der Rückschau erneut betrachtet werden können. Alles, was erledigt ist (das meiste ist erledigt oder nicht mehr relevant), ist eher zufällig erledigt worden. Für die neue Roadmap wird angeregt, sie in Jira anzulegen. Auf dem deutschen Anwendertreffen wurde ein Anfang dafür gemacht.

Bei einer Breakout-Session-Arbeitsgruppe, die sich mit dem Aufsetzen von VuFind 2.0 beschäftigt, stellt sich heraus, dass es noch Punkte gibt, die bei der Installation Schwierigkeiten machen.

Demian stellt kurz den Sinn und die Umsetzung von Unit Tests vor. Er ermuntert alle, Tests zu schreiben und damit die Codeabdeckung von 8,2 % zu erhöhen. Er bezeichnet die Tests als guten Startpunkt, sich mit Vufind 2.0 auseinanderzusetzen. Ein Testsystem steht unter <http://vufind.org/hudson> zur Verfügung. Die Tests liegen im VuFind git unter `modules/VuFind/tests/unit-tests`.

Zuguterletzt fordert Demian auch alle auf, die Votingfunktionen in Jira vermehrt zu nutzen, um eine bessere Priorisierung zu ermöglichen. Es wird der Vorschlag gemacht, einen wöchentlichen oder monatlichen „Voting Day“ einzurichten, um das Voting zu pushen.

Verschiedenes

In einer weiteren Breakout-Session wird die Frage behandelt, wie Vufind in einem Konsortium betrieben werden kann. Einige Teilnehmer agieren trotz vorhandener Schwierigkeiten mit aktuellen Verfügbarkeitsdaten, auf die über Facetten zugegriffen werden kann, die also indiziert sind.

Filipe MS Bento von der University Aveiro, Portugal, stellt in diesem Zusammenhang sein Konzept des „Searching 4.0“ vor. Unter Searching 4.0 versteht Filipe die Einbeziehung von usergenerierten (z.B. Tags), aber auch nicht usergenerierten Metadaten (z.B. Umlaufdaten). Diese Beziehungen können auch für die Suche benutzt werden. Filipe präsentiert auch Möglichkeiten zur Einbindung dieser zusätzlichen, sozialen Daten in die VuFind-Oberfläche vor. Er zeigt die Einbindung von Suchvorschlägen, „Ähnliche Suchanfragen“, „Ähnliche Autoren“, „Ähnliche Ressourcen“ sowie Browsing. Zudem geht Filipe darauf ein, wie Soziale Netzwerke wie Facebook in VuFind integriert werden können (so dass zum Beispiel ein Like-Button bei einer Ressource angezeigt wird) oder wie man Ergebnisse aus VuFind in andere soziale Bookmarkingdienste exportieren kann.

Mohan Raj Pradhan berichtet von healthnet Nepal, wo eine Kombination aus Koha, VuFind und Pazpar2 eingesetzt wird. Zusätzliche Datenbanken werden nur über Z39.50 eingebunden, wobei healthnet Hilfe aus Villanova bekam. Pazpar2 dient als freie Alternative zu teuren Indizes wie Summon, Ebsco oder Primo Central.

Da in Südostasien viele verschiedene Zeichensätze genutzt werden, die alle in einem gemeinsamen System unter einen Hut gebracht werden müssen, ist die Benutzung von Unicode besonders wichtig.

Chris Hallberg, der von Seiten der Villanova University healthnet unterstützt hat, berichtet von der einzigen Schwierigkeit, die sich stellte: die Sessions in Pazpar2 sind äußerst kurzlebig (eine Minute), so dass ein Parsen der Daten nicht so einfach möglich ist.

Mark Noble vom Marmot Library Network (USA) berichtet über die Einbindung von eContent in VuFind. Er referiert über die Einbindung aus einem System namens Overdrive, das nur unzureichend Schnittstellen besitzt und daher (im Bereich der Ausleihe von eBooks und der Nutzerdaten) per Screenscraping genutzt wird.

Im Index sind auch ebrary Inhalte vorhanden. Die Inhalte werden über einen selbst geschriebenen RecordDriver (econtentRecord) angezeigt. Die Indizierung der Inhalte findet über eine API von Overdrive statt. Elektronische Inhalte können mit dem eigenen Treiber direkt in VuFind abgebildet

werden, so dass die eBooks auch direkt ausgeliehen werden können. Die Ausleihe der Bücher läuft in Overdrive nach sieben Tagen automatisch von selbst aus.

Code4Lib Mid-Atlantic, 17.10.2012

Das Code4Lib Mid-Atlantic-Meeting ist das erste in dieser Form. Bei Code4Lib handelt es sich um eine englischsprachige Mailingliste, über die ein Austausch zum Thema „Software in Bibliotheken“ bzw. „Softwareentwicklung in Bibliotheken“ diskutiert wird. **David Uspall** von der Villanova University eröffnet diese eintägige Konferenz.

Erster Sprecher ist **John Ockerbloom** von der University of Pennsylvania, der die Global Open Knowledge Base vorstellt. Es handelt sich dabei um ein Projekt, an dem jeder teilnehmen kann (KB+). Ziele des Projektes sind die Entwicklung eines ERM-Modells, der Anschluss eines ERM-Repositories sowie die Entwicklung offener Schnittstellen dafür. Ähnlich wie beim Quali OLE-Projekt handelt es sich um eine kooperative Entwicklung mehrerer Bildungseinrichtungen. Das Projekt startete im Juni 2012 und soll bis Juni 2014 laufen.

Das Modell sieht vor, dass jede Entität im Repository einen global eindeutigen Identifier erhält. Auch eine Beziehungsbildung zwischen Entities soll möglich sein.

Um das Repository zu bevölkern, ist es notwendig, die Daten zunächst zu sammeln, über verschiedene (quellenabhängige) Wege. Die Daten müssen anschließend normalisiert (also in ein gemeinsames Format überführt) und bereinigt werden (z.B. Duplikatentfernung). Die Verteilung der Daten erfolgt dann wieder über unterschiedliche Wege: APIs, Websites, Downloads, ... GoKb soll dabei LOD-kompatibel sein bzw. auf LOD basieren.

Das Projekt ist erweiterbar und John stellt einige denkbare Erweiterungen vor:

- Ergänzung von Lizenzinformationen bzw. Open-Access-Informationen zu den Entities
- Erfassung von Inhaltsverzeichnissen zu Zeitschriften
- Archivierungsmöglichkeiten (Self-Archiving) für Autoren
- Dauerarchivierung

Derzeit ist es nicht geplant, die Daten auf Artikelebene aufzunehmen, auch das wäre eine denkbare Ergänzung. John fordert dazu auf, sich mögliche Einsatzszenarien für die GoKb zu überlegen. Nähere Informationen zu gokb werden auf der gokb.org Webseite bereitgestellt.

Als nächste beschreibt **Katherine Lynch** von der Drexel University, wie ein Mitarbeiter in Web Accessibility bzw. HTML und CSS eingeführt wurde. Der Mitarbeiter war bereits gewohnt, C++ zu programmieren, aber hatte noch keine Erfahrungen mit Webdevelopment. Katherine beschreibt, dass er auf Mac arbeitete und als Entwicklungsumgebung mit Smultron sowie mit Cyberduck als FTP-Client gearbeitet hat. Um die Web-Accessibility einer Website zu testen, können verschiedene Methoden angewendet werden. Um zu sehen, wie Blinde die Webseite sehen, kann zum Beispiel der Monitor ausgeschaltet und ein Screenreader angeschaltet werden, der die Webseite vorliest. Es sollte auch getestet werden, ob alle Elemente der Website per Tastatur zugänglich sind.

Die Reihenfolge im Sourcecode ist sehr wichtig, denn die Elemente werden in genau dieser Reihenfolge vorgelesen (CSS deaktivieren: ist der Inhalt der Seite immer noch verständlich?).

Wichtige Elemente sollten möglichst weit oben angeordnet werden. Katherine weist auch auf die Wichtigkeit von alt-Texten für Bilder hin.

Für Windows existieren mehrere, frei verfügbare Screenreader: Thunder, NVDA oder Jaws. Ein sinnvoller Test, um die Zugänglichkeit per Tastatur zu testen, ist es auch, die Maus herauszuziehen. Durch derartige Tests kann die Zugänglichkeit der Webseite ermittelt und gezielt verbessert werden. Auch ein Feedback-Formular kann helfen, damit Menschen, die auf Probleme stoßen, ihr Feedback sofort einsenden können. Weiteres zum Thema findet sich unter www.katherinelynch.org.

Mohan Raj Pradhan von healthnet Nepal, der auch schon auf dem VuFind Summit referiert hatte,

berichtet erneut über die verschiedenen Komponenten, mit denen VuFind in Nepal kombiniert wurde. Zudem wurde in Nepal eine Live-DVD für Koha erstellt, mit der ein laufendes Koha-System ohne Installation gestartet werden kann. Die DVD kann im Netz heruntergeladen werden. Mohan stellt zwei Komponenten zur Verwaltung elektronischer Ressourcen vor: CUFTS und GODOT. CUFTS kann als Alternative zu kommerziellen Indizes verwendet und eine Knowledge Base bereitgestellt werden. Das Programm ist Perl-basiert und muss lokal installiert und eingerichtet werden. GODOT dient als Link-Resolver, der OpenURLs für die elektronischen Ressourcen erstellt. Über CJDB (eine Komponente von CUFTS) ist eine Verwaltung der abonnierten elektronischen Ressourcen möglich.

Als nächste Rednerin stellt **Michelle Suranofsky** von der Lehigh University die Arbeit am Kuali OLE Projekt vor. Bei Kuali OLE soll ein vollständiges Bibliothekssystem erstellt werden. Auch die Villanova University nimmt am Kuali-Netzwerk teil. Es ist ein sehr großes Projekt mit vielen Partnern. Zur Beschreibung von Funktionalitäten wurden User Stories erstellt. Kuali entsteht auf einer Plattform namens KRAD und bedient sich wie VuFind einer Solr-Suchmaschine. Die Universitäten von Chicago und Lehigh sollen die ersten Anwender von Kuali OLE werden. Geplant ist der Produktionseinsatz für den Sommer 2013.

Karen Coombs von OCLC vergleicht objektorientierte Programmierung und Rest-APIs miteinander. Sie konstatiert, dass OCLC in der Regel REST-Schnittstellen mit der Software bereitstellt. REST-Schnittstellen ermöglichen eine plattformunabhängige Weiterverarbeitung der Daten.

Die von ihr beschriebenen Schnittstellen sollen in OCLCs Produkt Web Scale Management (WMS) enthalten sein. In der jetzigen Version des ebenfalls von OCLC hergestellten Bibliothekssystems PICA, das an der TUBHH (und im gesamten GBV) eingesetzt wird, fehlen derartige Schnittstellen fast vollständig.

VuStuff III, 18.10.2012

Nach der Einführung durch **David Uspall** berichtet **Rebecca Goldman** über das Thema Langzeitarchivierung. Sie zeigt zunächst ein Buch, in dem noch 5 ¼ Zoll-Disketten enthalten sind. Damit verbunden ist die Frage, wie viele Bibliotheken tatsächlich noch in der Lage sind, diese Medien zu lesen. Auch die Systemvoraussetzungen für die Software auf den Disketten sind 30 Jahre alt und demzufolge auf heutige Systeme nicht mehr übertragbar (MS DOS 2.0...). Rebecca berichtet, dass die erwartete Nutzbarkeit digitaler Datenträger bei maximal 15 Jahren liegt. Digitalisierung allein bedeutet auch keine Langzeitarchivierung, da Formate evtl. später auch nicht mehr gelesen werden können. Auch Ausdrucken ist kein Allheilmittel, weil viel zu viel Platz verbraucht werden würde, sich einige Inhalte gar nicht ausdrucken lassen und die Inhalte außerdem nach dem Ausdruck nur noch schwer auffindbar sind (keine Suchfunktionalität mehr möglich). Mit Dateien müssen laut Rebecca immer auch Metainformationen abgespeichert und archiviert werden, damit klar ist, wie die Dateien behandelt werden müssen.

Rebecca definiert die Langzeitarchivierung (LZA) allgemein als dauerhafte Zugänglichmachung digitaler Materialien. Strategien dafür sind:

- Backups
- Verteilte Datenhaltung
- Checksummenberechnung (Nachweis der Originalität und Authentizität der Datei, so kann sichergestellt werden, dass die Datei in ihrem ursprünglichen Zustand vorliegt und nicht verändert wurde)
- Migration auf neue Plattformen
- Emulation alter Plattformen
- Aufbewahrung der Hardware

Eine wichtige Frage bei elektronischen Ressourcen ist, ob sie wirklich besessen oder nur lizenziert werden. Bei lizenzierten Materialien, die im Originalzustand nicht auf dem eigenen Rechner sind, ist eine Langzeitarchivierung gar nicht möglich.

Rebecca weist darauf hin, dass zur Durchführung von LZA auch Personal und Geld benötigt werden. Als Beispiel zur Langzeitarchivierung von Webseiten weist Rebecca auf archive-it.org hin. Archive-it ist im Gegensatz zu [internetarchive](http://internetarchive.org) durchsuchbar. DuraCloud ist ein weiteres LZA-Projekt.

Stephen Spatz erläutert die Erstellung eines digitalen Musikarchivs von irischer Musik.

Liveaufnahmen (auf Tape vorliegend) der Ceili Group sollen digitalisiert werden. Dabei mussten zunächst technische Fragen (wie Formatfragen) geklärt werden. Früher wurde die CD als perfektes Medium zur Langzeitarchivierung musikalischer Master angesehen. Das hat sich heute geändert; die Qualität reicht zur Originalarchivierung nicht mehr aus.

Die Digitalisierung der Originalaufnahmen, die auf Tape vorliegen, erfolgt mit Hilfe von Audacity. Es wird immer der wav-Master und die komprimierte mp3-Version archiviert. Jede Tapeseite wird zu einem Master verarbeitet, der anschließend geschnitten und in Tracks aufgesplittet wird.

Gesprochene Passagen (und Lyrics?) werden neben der Musikdatei auch als doc-Dokument über die Plattform bereitgestellt. Auf den Tapes befinden sich immer wieder Stör- und Rauschgeräusche, die nur schwer zu entfernen sind. Das Stimmen der Instrumente und ähnliche unbrauchbare Stellen werden aus der geschnittenen Fassung entfernt.

Es gibt einen Preview auf Emilie Davies Tagebuch. Es handelt sich um ein Tagebuch einer Frau, die in der Zeit der amerikanischen Unabhängigkeitserklärung lebte und Notizen zum Tagesgeschehen machte – ein nahezu einzigartiges historisches Dokument. Dieses Tagebuch wurde an der Villanova University digitalisiert und aufbereitet. Die Ergebnisse werden über eine Wordpress-Plattform bereitgestellt. Dabei hat eine Wordpress-Seite 3 bis 4 Tageseinträge als Inhalt. Die Inhalte sind transkribiert, aber auf der Seite ist auch der Originalscan verfügbar. Über Kommentare können

Korrekturvorschläge bzw. Interpretationen von jedem Interessenten bereitgestellt werden. Auch von den Autorinnen des Projektes werden Annotationen angefertigt. Das Original ist aufgrund der alten Schrift und altersbedingter Abnutzungsspuren großteils nur schwer zu entziffern. Die Ergebnisse finden sich unter <http://davididiaries.villanova.edu>, der Inhalt wird Anfang 2013 freigeschaltet.

Auch das Thema RDA und LOD steht auf dem Programm. In diesem Zusammenhang wird unter anderem auf das Problem der Homonyme im Netz hingewiesen. Es ist zum Beispiel mehrdeutig, welches Boston gemeint ist, da es viele Boston auf der Welt gibt. Auch bei Worten wie Apple ist nicht klar, ob die Frucht oder die Firma gemeint ist. RDA beschreibt daher Entitäten, und nicht Worte. RDA basiert dabei auf FRBR. Über einen URI (Unified Resource Identifier) wird eine Entität eindeutig gekennzeichnet.

Im klassischen RDF-Schema besteht eine Entität aus einem Triplet: Subjekt, Prädikat und Objekt. Subjekt und Prädikat müssen dabei immer URIs sein, das Objekt kann entweder eine URI oder ein Text sein. Das Subjekt muss auch aus einer Ontologie bestehen. Ontologien können dabei beliebig kombiniert und wiederverwendet werden.

Durch die Beziehungsbildung zwischen Entitäten können auch Rückschlüsse gezogen haben. Zum Beispiel lässt sich aus der Charakterisierung „Big Ben“ ist in „London“, „London“ ist in „Großbritannien“ schließen, dass auch „Big Ben“ ist in „Großbritannien“ zutrifft.

Demian Katz berichtet zum Abschluss über die Digitalisierung für das Projekt Gutenberg. Um zum Projekt Gutenberg beizutragen, muss man zunächst dafür sorgen, dass eine funktionierende OCR-Software vorliegt. Dazu ist es notwendig, die Software zu testen, indem man die erkannten Texte korrekturliest. Zudem muss überprüft werden, ob das Layout korrekt erkannt wurde und den Inhalt nicht durcheinanderbringt.

Anschließend muss ein Buch ausgewählt werden, das man beisteuern darf (gemeinfrei), das aber noch nicht im Projekt Gutenberg vorhanden ist.

Die Scans werden dann auf pgdp.net hochgeladen und können dort verteilt korrekturgelesen werden. Vor der Freigabe muss nochmal sichergestellt werden, dass alles korrekt ist. Das Korrekturlesen kann auf pgdp.net auch anderen überlassen werden.

Den Preis für den besten Vortrag erhält nach Abstimmung des Publikums Rebecca Goldman.

Fazit

Insgesamt hat mir persönlich die Teilnahme an der Tech-Trifecta-Woche sehr viel gebracht. Ich konnte viele neue Kontakte schließen, und die Personen, mit denen ich bislang nur virtuell zusammengearbeitet habe, auch persönlich im direkten Kontakt kennenlernen. Insbesondere möchte ich dabei die Kontakte zu Demian Katz und Filipe Bento erwähnen, aber auch viele weitere neue Kontakte konnten geschlossen und mein Netzwerk damit beträchtlich und sinnvoll erweitert werden.

Durch die Teilnahme an allen drei Konferenzen konnte ein sehr breites Themenspektrum abgedeckt werden, das weit über den Einsatz von VuFind hinaus ging. So konnte ich sehr viele unterschiedliche und neue Ansätze für die verschiedenen Themen von der Reise mitnehmen. Ich habe mich bemüht, die verschiedenen Aspekte in dieser Zusammenfassung darzustellen und somit nicht nur für mich den Inhalt der Tagungen zu dokumentieren.

Ich danke BI International sehr herzlich dafür, mir diese Reise ermöglicht zu haben!