Richtlinie zum „Software-Sprint“

"Documents4Democracy"

Schlussbericht

Zuwendungsempfänger:

Svante Schubert

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS17S36 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Was war Deine Motivation? Welches Problem wolltest Du mit Deinem Projekt lösen? Wie war die geplante Vorgehensweise zur Problemlösung (auch Angabe der wichtigsten Meilensteine)?

Die technische Problematik die dem Prototyp zu Grunde liegt, resultiert aus dem Bedürfnis mit vielen Menschen gemeinsam an Dokumenten zu arbeiten. Technisch gesehen fehlt uns noch immer diese Freiheit, mittels unterschiedlicher Anwendungen zeitgleich und selbst offline, gemeinsam an einem Dokument zu arbeiten. Die kollaborativen Fähigkeiten von Etherpad, Google Docs, Office 365, etc. werden nur intern geboten. Darüber hinaus gibt es für komplexe (Office) Dokumente nicht die Möglichkeit, Änderungsvorschläge zu verschicken - ähnlich einem 'PullRequest' von Github. Auch eine lückenlose Historie, sowie Ansicht von Änderungen ab einem bestimmten Zeitpunkt (z.B. seit Urlaubsabwesenheit) sind nicht anzeigbar. Diese Aufzählung könnte beliebig weitergeführt werden.

Meilensteine

1. Erweiterungen des Apache ODF Toolkit Projekt (incubating) um folgende Basisfunktionalitäten:

   Anhand von ODF XML SAX Events werden logische Dokumentobjekte identifiziert, wie sie der Benutzer häufig in Dokumenten beim Einfügen, Löschen und Ändern kennt. Diese Änderungen werden dann zu dem im Web üblichen JSON Format transformiert, um später mit anderen ODF Office Anwendungen ausgetauscht zu werden.
   Grundsätzlich kann ein ODT Dokument in die Bibliothek eingegeben und als eine Liste von Änderungen der im Folgenden genannten Feature ausgegeben werden. Beliebige neue Änderungen können im Anschluss auf das Dokument angewendet werden. Dies alles erfolgt auf der Basis von der neuen Dokumentänderungen Basistechnologie.

2. Neues Test Tooling für das Regression Testen von Änderungen sind erforderlich.
    a. Ein Dokument wird in eine Liste mit Änderungen überführt und dann mit einer Referenzliste verglichen.
    b. Die Liste mit Änderungen wird auf ein leeres Dokument ausgeführt und danach wieder ausgelesen, um erneut mit einer zuvor gesicherten Referenzänderungsliste verglichen zu werden.
3. Für folgende Feature müssen Testdokumente zur Verfügung gestellt und Tests erfolgreich abgeschlossen werden:
    a. Implementierung von Änderungen für Paragraphen und Text
    b. Implementierung von Änderungen von Styles
    c. Implementierung von Änderungen für Listen
    d. Implementierung von Änderungen für Links
    e. Implementierung von Änderungen für Tabellen, Spalten, Reihen und Zellen

Beitrag des Projektes zu den Zielen der Förderinitative „Software-Sprint“

Welche Bezüge gibt es zu den Themenfeldern „Civic Tech“ und „Data Literacy“ des Software Sprints oder zu weiteren gesellschaftlich relevanten Zielen bzw. Lösungsansätzen?

Der Bezug zu den Themenfeldern „Civic Tech“ und „Data Literacy“ des Software Sprints besteht darin, dass ein Prototyp zur Basistechnologie der gemeinsamen Arbeit an komplexen (Office-) Dokumenten erstellt wurde. Der Prototyp vermag ein Dokument in seine Benutzeränderungen zu zerlegen, die als Reihe ausgegeben werden. Auch neue Änderungen (durch andere Nutzer) können bereits, z.B. via Position, in das Dokument übertragen und zusammengeführt werden.

Diese Technologie ermöglicht Office Anwendungen Änderungen, anstatt wie bisher lediglich Dokumente auszutauschen. Es ermöglicht gleichzeitig auch offline mit anderen Nutzern am selben Dokument zu arbeiten und Änderungen im Anschluß im Dokument zusammenzuführen.

Durch die gleichzeitige internationale Standardisierung der Änderungen bei OASIS (und im Anschluß bei der ISO) wird auch eine hohe Interoperabilität zwischen Anwendungen gewährleistet.

Ausführliche Darstellung der Ergebnisse

Welche konkreten Ergebnisse hast Du erzielt? Konnten alle Meilensteine erreicht werden? Welche zusätzlichen Erkenntnisse hast Du aus der Projektarbeit gewonnen, auch im Hinblick auf die Begleitung durch die Open Knowledge Foundation?

Mit etwas Mühe konnten letztendlich alle Meilensteine (s.o.) heute am 21.04.2018 erreicht werden und als Implementierung unter GitHub [1,2] bereitgestellt werden.

[1] https://github.com/svanteschubert/odftoolkit/tree/odf-changes
[2] https://github.com/svanteschubert/odftoolkit/tree/code-generation

In den folgenden Wochen bis zur CeBIT Präsentation wird die Code Basis noch refaktoriert und letzte Fehler behoben werden. Vor allem wird die Dokumentation verbessert, die bislang im Wesentlichen lediglich im Code zu finden ist.
Folgende Dokumentationen stehen noch vor der Veröffentlichung:

1. Einführende Erläuterung der Domain Problematik
2. Idee der Problemlösung aus der Vogelperspektive
3. Architekturbeschreibung des Prototypes basierend auf dem Apache ODF Toolkit Projekt
4. Beschreibung der ODF Änderungen
5. Beispiele und die Möglichkeit eigene Office Dateien in Änderungen zu zerlegen

Dazu ist auch die Zusammenführung des Prototypes mit dem Apache ODFToolkit (incubating) geplant, doch der Wechsel von Subversion zu Git(hub) verzögerte sich bei Apache, so daß das offizielle Github bis heute nicht zum Schreiben freigeschaltet ist.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Welcher Nutzen ergibt sich für die Zielgruppe aus den Ergebnissen Deines Projekts? Welche weitergehenden Effekte ergeben sich aus der Open-Source-Stellung der Ergebnisse? Gibt es Ideen für die Weiterentwicklung Deiner Lösung und Pläne zu deren Umsetzung?

Ideen zur Weiterentwicklung gibt es viele (s.u. Kurze Darstellung der Aufgabenstellung und Motiviation). Besonders das Einbeziehen eines vollzeitlichen LibreOffice Entwicklers in diese Thematik, wahrscheinlich beeinflusst durch meinen Vortrag über dieses Projekt auf der FOSDEM im Februar diesen Jahres, ist ein großer Erfolg.

Die Zielgruppe (ODF Office Anwendungen) wird über weitere Konferenzen (z.B. LibreOffice Konferenz 2018) und der OASIS OpenDocument Arbeitsgruppe über die Weiterentwicklung auf dem Laufenden gehalten.

Hat die Arbeit in dem Projekt Dich in Deiner persönlichen, fachlichen Weiterentwicklung unterstützt?

Abgesehen von der üblichen Unterschätzung des Zeitrahmens, der am Ende dennoch gemeistert wurde, waren die begleitenden Fortbildungen über den PrototypFund inspirierend und fördernd.

Ich denke ohne die Förderung, hätte ich diesen riesigen Kraftaufwand für das OpenSource Projekt nicht ohne weiteres auf mich genommen. Danke! 😊

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Gab es Arbeiten bzw. Lösungsansätze, die nicht weiter verfolgt wurden? Was waren die Hintergründe, und wie bist Du alternativ vorgegangen?

Anfangs war geplant weitaus mehr Source Code aus der XML Grammatik des OpenDokument Formates zu generieren. Dabei wurde das XML der ODF Grammatik über ANTLR (einem Parser Generator) in GraphML Dateien überführt und diese in die Apache Tinkerpop Graphdatenbank geladen. Doch die Einarbeitung in die Scriptsprache Gremlin erwies sich als schwierig und die notwendige Geschwindigkeit konnte nicht gehalten werden.

Daher wurde ein anderer manueller Ansatz gewählt, der dadurch weitaus mehr manuelle Programmierung erforderte. Es ist jedoch geplant in der Zukunft die Generierung zu Perfektionieren und den manuell erstellten Code gegen generierten Code schrittweise und ohne Funktionsverlust auszutauschen, bzw. zu refaktorieren.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Wo können sich Interessenten detailliert über Deine Projektergebnisse informieren (z.B. Webseite, GHitHub, Veröffentlichungen)?

Neben der bereits erwähnten beiden GitHub Branches (s.o), die noch in den Hauptzweig des Apache ODF Toolkit eingearbeitet werden müssen, steht wie erwähnt ein Großteil der Dokumentation und Beispiele noch aus.

Es wurde bereits ein Vortrag auf der OpenSource Konferenz FOSDEM in Belgien über die Arbeit gehalten und neben der Präsentation für den PrototypeFund auf der CeBIT ist auch eine Präsentation auf der LibreOffice Konferenz in Albanien geplant. Auch ein langjähriger LibreOffice Entwickler hat sich dem Thema bei LibreOffice angenommen.

Weitere Dokumentation über ODF Änderungen wird im OASIS Subcommittee für „Advanced Document Collaboration“ zu erwarten sein, welches ich leite und in welchem die ODF Änderungen standardisiert werden sollen.

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Gab es im Projektverlauf Ereignisse, die eine Anpassung der Planung erforderlich machten – z.B. Mehr- oder Minderaufwand bei der Bearbeitung von Teilaufgaben?

Ja, der Aufwand bei der Einarbeitung in die Graphendatenbanken war wesentlich schwieriger als angenommen und wurde verschoben.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Gab es Entwicklungen anderer Personen oder Institutionen, die Einfluss auf Deine Arbeiten und die Zielsetzung hatten? Wenn ja, worin bestand dieser und wie bist Du damit umgegangen?

Auf Konferenzen und auch mit Interessierten wurden lange Diskussionen über das Thema geführt, die in die Arbeit des Projektes mit einflossen.