

Von und nach L^AT_EX mit Pandoc

Blandyna Bogdol
blandyna.bogdol@gmail.com

02. November 2013

Über die Person

- Studium der Elektrotechnik
- 10 Jahre Berufserfahrung als technische Redakteurin
- Arbeitsschwerpunkt: Softwaredokumentation
- Arbeitswerkzeuge: Open Source (nach Möglichkeit)
- Immer wieder im \LaTeX -Umfeld unterwegs
- Hobbys: Fotografie, Tanzen als Leistungssport

Heutiges Thema: Pandoc

- Brauchen wir einen Konvertierer?
- Einsatzszenario
- Kurze Einführung in Pandoc
- Ein paar kleine Beispiele von Pandoc

Allgemeine Gedanken zum heutigen Thema

- Der Einsatz von Pandoc mit praktischen Beispielen wurde schon ausführlich beschrieben (vor allem mit Markdown)
- Es soll keine Einführung in die Konfiguration von Pandoc sein
- Allgemeine Gedanken über Sinn des Konverters
- Überlegungen zu Vorteilen und Grenzen eines Konverters

Brauchen wir einen Konvertierer in \LaTeX -Umfeld?

- \LaTeX als einziges Format reicht nicht immer aus (kein alltägliches Format, komplexe Datenstrukturen wie Handlungsanweisungen nicht möglich...)
- Einstieg in \LaTeX ist nicht ganz einfach
- Viele verschiedene Daten und Dokumente in unterschiedlichen Formaten
- Übersetzbarkeit mit Translation Memory Tools nur schwer zu gewährleisten (eigene Konverter für jedes Werkzeug nötig)

- Arbeitsumfeld mit unterschiedlichen Eingabeformaten (RTF, ODT, HTML, Markdown, XML...)
- Ausgabeformate: hochwertiges PDF, HTML und "Korrekturformat" (ODT, RTF oder DOC)
- Ein Werkzeug für alle Umwandlungsschritte (manuelle Formatierung soll entfallen)

Was ist Pandoc

- Pandoc = universeller Dokumentformatkonvertierer ("Multi target publishing")
- Basiert auf der Programmiersprache Haskell (rein funktionale Sprache)
- Plattformunabhängig
- Einfache Installation
- Umfangreiche Dokumentation auf der Webseite (<http://johnmacfarlane.net/pandoc/index.html>)
- Viele freie Add-Ons

Funktionsweise von Pandoc

- Kommandozeilentool
- Reader und Writer (es gibt mehr Writer als Reader)
- Konfigurationsdateien können / müssen manuell angepasst werden
- Einlesen von ODT, RTF und DOC nicht möglich
- \LaTeX -Engines: \LaTeX , XeLaTeX, LuaLaTeX und ConTeXt

Installation von Pandoc (Linux-Systeme)

- Haskell über Paketverwaltung installieren
- Auf der Konsole folgende Befehle ausführen:
`cabal update`
`cabal install pandoc`
- Konfigurationsdateien:
`usr/share/pandoc-Version/templates`

Pandoc verwenden

- Befehl (allg. gültig):

```
pandoc -t ReaderFormat -f WriterFormat  
--template=Datei Eingabe.xyz -o  
Ausgabe.xyz
```

- Eine Vielzahl weiterer Parameter möglich (siehe Dokumentation)!
- Ausgangsdokumente müssen konsistent sein (Redaktionsleitfaden)

Beispiel: Markdown nach L^AT_EX

- Beispieldatei: README
- L^AT_EX: `pandoc -s README -o Readme.tex`
- L^AT_EX mit Definition der Vorlage:
`pandoc --template=default.latex README -o test.tex`

Beispiel: Markdown nach PDF

- Beispieldatei: README
- PDF: `pandoc -s README -o Readme.pdf`
- PDF mit XeLaTeX: `pandoc README -o Readme-xelatex.pdf --latex-engine=xelatex`
- PDF mit XeLaTeX und Inhaltsverzeichnis:
`pandoc --toc README -o Readme-xelatex.pdf --latex-engine=xelatex`

- Pandoc funktioniert sehr gut für umfangreiche Dokumente, mit recht einfacher Struktur und ohne spezielle Verarbeitungsanweisungen
- Einfach heißt hier:
 - Viel Text, klare Struktur der Überschriften
 - Listen, Tabellen, Quellcodes
 - Keine semantischen Auszeichnungen wie Produktnamen, Menüreihenfolgen, Handlungsanweisungen
 - Keine Anweisungen für spezielle Weiterverarbeitung (Anweisungen für Übersetzungsprogramme etc.)

Von DocBook nach L^AT_EX

- Pandoc funktioniert mit DocBook 4.x
- Nicht alle Elemente von DocBook 4.x wurden implementiert
- DocBook 5.x (aktuelle Version) wird offiziell noch nicht unterstützt
- Einige Dokumente in DocBook 5.x lassen sich jedoch nach L^AT_EX umwandeln
- Komplexe Strukturen sind problematisch

- Ein Beispiel aus DocBook 5.x:
Klicken Sie auf das Menü Datei / Speichern.
- Quellcode:

Klicken Sie auf das Menü

```
<menuchoice>
```

```
<guimenu>Datei</guimenu>
```

```
<guisubmenu>Speichern</guisubmenu>
```

```
</menuchoice>.
```

Von DocBook nach L^AT_EX – 3

- Ein Beispiel aus DocBook 5.x: "*Klicken Sie auf das Menü Datei.*"
- Quellcode:

Klicken Sie auf das Menü

```
<menuchoice>
```

```
<guimenu>Datei</guimenu>
```

```
</menuchoice>.
```

Klicken Sie auf das Menü

```
<guimenu>Datei</guimenu>.
```

- Das Spiel kann mit Tastenkürzel für Menüs viel weiter getrieben werden!

- Befehl für Pandoc:

```
pandoc -f docbook -t latex  
--template=default.latex Eingabe.xml -o  
Ausgabe.tex
```

- **Achtung:** Ohne die Angabe der Vorlage ist die Ausgabedatei nicht vollständig!

Von L^AT_EX nach DocBook

- Hauptgründe für DocBook: Industriestandard und Weiterverarbeitungsmöglichkeiten
- Befehl für Pandoc (Ausgabe nur als DocBook 4.4):

```
pandoc -f latex -t docbook  
--template=default.docbook Eingabe.tex -o  
Ausgabe.xml
```
- **Achtung:** Ohne die Angabe der Vorlage ist die Ausgabedatei nicht vollständig!

Zusammenfassung: Vorteile

- Pandoc ist ein sehr vielseitiger Konvertierer
- Verfügbar für alle Betriebssysteme und OpenSource
- Sehr umfangreiche Dokumentation
- Eigene Anpassungen und Add-Ons sind möglich
- Sehr gut für Dokumente mit einfacher Struktur (Textmenge spielt keine Rolle)
- Schnell und vielseitig einsetzbar

Zusammenfassung: Nachteile

- Komplexe Strukturen von Dokumenten werden nicht richtig erkannt
- Anpassung der Vorlagen ist aufwendig
- Wenig aussagekräftige Fehlermeldungen
- Nicht klar ersichtlich, wann Angaben zum Reader / Writer und den Vorlagen gemacht werden müssen
- Einlesen von Word-Dateien ist nicht möglich
- Vorlagen-Dokumentation könnte umfangreicher sein

- Sehr guter Konverter, der ständig weiter entwickelt wird
- Empfohlen bei vielen, gleichen Dokumenten mit einfacher Struktur
- Gute Möglichkeiten, viele Formate mit einem Werkzeug zu erzeugen

- 1 Axel Kielhorn "Viele Ziele Veröffentlichung" (Stand: 2013.02.03),
<https://github.com/AKielhorn/Markdown-Intro/blob/master/Ziele-md.md>
- 2 Daniel Stender "PDF-Dokumente schreiben mit Pandoc und Markdown", freiesMagazin 06/2013, Seiten 12–16
- 3 "Pandoc and foreign characters" (Quelle: <https://stackoverflow.com>)
- 4 Pandoc-Webseite:
<http://johnmacfarlane.net/pandoc/index.html>