

Entwicklung eines prototypischen Konverters von Single-Markdown zu Multi-HTML

- Modul Crossmedia-Produktion
- Philipp Anders, Ferdinand Malcher
- MIM 15

Markdown ist eine leichtgewichtige Auszeichnungssprache für strukturierte Dokumente. Zu Präsentationszwecken findet eine Konvertierung in verschiedene Dokumentenformate wie z.B. DOCX, LaTeX, Docbook, TXT, etc. statt. Die Umwandlung zu HTML hat sich hierbei als De-Facto-Standard etabliert. Durch die minimalistische Syntax sind Markdown-Dokumente auch vor deren Umwandlung vom Menschen leicht lesbar. Die reduzierten Ausdrucksmittel bieten auch Einsteigern einen intuitiven Zugang zur Dokumentenbeschreibung.

Bei der Konvertierung von Markdown zu HTML arbeiten die konventionellen Konvertierungstools üblicherweise *Single-Markdown to Single-HTML*, d.h. sie überführen *eine* Markdown-Datei in genau *eine* HTML-Seite. Dem gegenüber existieren spezialisierte Tools zur *Multi-Markdown to Single-HTML*-Konvertierung für umfangreiche Projekte, d.h. die Übertragung einer Kollektion von Markdown-Files in eine HTML-Seite mit menüartiger Navigation (ähnlich einem Inhaltsverzeichnis).

Beispiele hierfür sind:

- **MKDocs:** <http://www.mkdocs.org/>
- **DAUX.io:** <http://daux.io/index>
- **gulp-markdown-docs:** <https://github.com/sojournerc/gulp-markdown-docs>

All diese Tools gehen allerdings davon aus, dass

- die zu kombinierenden Markdown-Quelldateien separat vorliegen
- zu jeder Quelldatei vom Autor Metadaten angegeben werden, die deren Kombination spezifizieren

Ziel des Projekt

Ziel dieses Projekts ist die Entwicklung eines prototypischen "Single-Markdown to Multi-HTML"-Konverters. Dieser soll *ein* Markdown-Dokument in mehrere, separate HTML-Seiten mit menüartiger Navigation umwandeln. **Separate HTML-Seiten ist hierbei im Sinne einzelner, durch eine URL identifizierbarer Ansichten gemeint, nicht zwangsläufig im Sinne separater HTML-Dateien.** Dabei soll die Aufspaltung auf den Dokumenten-Abschnitten basieren, die implizit durch die Überschriften gebildet werden.

Anforderungen:

1. Zerlegung: Eine HTML-Seite soll einen Abschnitt erster Ebene beinhalten.
2. Navigation: Die Navigation zwischen den HTML-Seiten soll über eine Menüleiste möglich sein. Diese soll Überschriften bis zur zweiten Ebene beinhalten.
3. Cross-Referenzierung: Alle Referenzen innerhalb des Dokuments (*Cross-Referenzen*) sollen erhalten bleiben.

Der Aufruf des Tools soll über die Kommandozeile erfolgen:

```
md2multihtml input.md output.html
```

Anwendung

Die Konvertierung in ein navigierbares, mehrseitiges HTML-Dokument eignet sich für prinzipiell für alle umfangreichen Dokumente mit komplexer Struktur. Die Nutzung nur einer einzigen Markdown-Quelldatei eignet sich besonders für kollaboratives Schreiben, wobei die Verwaltung mehrerer Quelldateien unpraktisch ist.

Konkrete Anwendungsfälle hierfür sind:

- Die Erstellung von Vorlesungsmitschriften. Hierbei wird über ein beliebiges Tool zum kollaborativen Schreiben das Skript gepflegt, zur Aufbereitung zum Lernen wird der Konverter genutzt.
- Die Erstellung von Softwaredokumentationen. Durch die Markdown-Unterstützung von Github hat sich Markdown besonders als Beschreibungsformat für Readme-Dateien etabliert. Umfangreiche Dokumentationen wie z.B. die von [MySQL](#) büßen durch die Darstellung in einer "langen" HTML-Seite jedoch an Übersichtlichkeit ein.

Technische Umsetzung

Die hierarchische Struktur des Dokuments muss analysiert und in eine baumartige Datenstruktur überführt werden. Zusätzlich müssen aufgrund der Cross-Referenzen im Dokument Knoten des Baumes mit zusätzlichen Kanten in Zusammenhang gebracht werden, um Verweise auch über Dokumententeile hinweg auflösen zu können. Auf dieser Art und Weise wird ein *Strukturgraph* des Dokuments erstellt. Dieser bildet die Basis der Zerlegung in separate HTML-Seiten, dem Aufbau der Navigation und der Auflösung von Cross-Referenzen.

Für die tatsächliche Umwandlung von Markdown zu HTML soll eine bestehende Bibliothek eingesetzt werden. Je nach Wahl des eingesetzten Markdown-Parsers sind ein direkter Eingriff in den erstellten Syntaxbaum oder eine Nachverarbeitung der generierten HTML-Datei denkbar.