

Pandoc & Markdown: Dream-Team zum Erzeugen von Multi-Channel Dokumenten-Output

Kurt Pfeifle,

22. August 2015
FrOSCon

- 1 Einführung
- 2 Pandoc
- 3 Zugaben!

Einführung

Markdown (I)

- Schlanke Syntax zum Markup von Text-Dokumenten
- “Erfinden” und spezifiziert 2004
- Inspiriert von “Email-Syntax aus vor-HTML-Zeiten”
- Urheber: John Gruber und
Aaron “*The Internet’s Own Boy*” Swartz (†)

Markdown (I)

- Schlanke Syntax zum Markup von Text-Dokumenten
- “Erfinden” und spezifiziert 2004
- Inspiriert von “Email-Syntax aus vor-HTML-Zeiten”
- Urheber: John Gruber und
Aaron “*The Internet’s Own Boy*” Swartz (†)

Tipps für die HTML-Version dieser Präsentation:

- 1 Eintippen von ? zeigt Hilfe (ESC zum Beenden).

Markdown (I)

- Schlanke Syntax zum Markup von Text-Dokumenten
- “Erfinden” und spezifiziert 2004
- Inspiriert von “Email-Syntax aus vor-HTML-Zeiten”
- Urheber: John Gruber und
Aaron “*The Internet’s Own Boy*” Swartz (†)

Tipps für die HTML-Version dieser Präsentation:

- 1 Eintippen von ? zeigt Hilfe (ESC zum Beenden).
- 2 Eintippen von f schaltet auf Vollbild.

Markdown (I)

- Schlanke Syntax zum Markup von Text-Dokumenten
- “Erfinden” und spezifiziert 2004
- Inspiriert von “Email-Syntax aus vor-HTML-Zeiten”
- Urheber: John Gruber und
Aaron “*The Internet’s Own Boy*” Swartz (†)

Tipps für die HTML-Version dieser Präsentation:

- 1 Eintippen von ? zeigt Hilfe (ESC zum Beenden).
- 2 Eintippen von f schaltet auf Vollbild.
- 3 Eintippen von o zeigt Überblick (ESC zum Beenden).

Markdown (I)

- Schlanke Syntax zum Markup von Text-Dokumenten
- “Erfinden” und spezifiziert 2004
- Inspiriert von “Email-Syntax aus vor-HTML-Zeiten”
- Urheber: John Gruber und
Aaron “*The Internet’s Own Boy*” Swartz (†)

Tipps für die HTML-Version dieser Präsentation:

- 1 Eintippen von ? zeigt Hilfe (ESC zum Beenden).
- 2 Eintippen von f schaltet auf Vollbild.
- 3 Eintippen von o zeigt Überblick (ESC zum Beenden).
- 4 Eintippen von b oder . pausiert die Präsentation.

Markdown (II)

Markdown (II)

- Ursprüngliches Ziel: Erzeugen von HTML aus Text
(*“für Leute, die kein HTML lernen möchten”*)

Markdown (II)

- Ursprüngliches Ziel: Erzeugen von HTML aus Text (*“für Leute, die kein HTML lernen möchten”*)
- Hierfür diente ein Perl-Skript, `"Markdown.PL"` *leider seit 2004 unverändert in Version 1.02!*

Markdown (II)

- Ursprüngliches Ziel: Erzeugen von HTML aus Text (*“für Leute, die kein HTML lernen möchten”*)
- Hierfür diente ein Perl-Skript, `"Markdown.PL"` *leider seit 2004 unverändert in Version 1.02!*
- Trotzdem überwältigender Erfolg für Markdown – weite Verbreitung: GitHub, StackExchange, Reddit, PHP, ...

Markdown (III)

- John Gruber zeigte sich störrisch, was den Einbau zusätzlicher Features betraf: es kommt daraufhin zur Entwicklung zahlreicher Erweiterungen vieler z.T. unvereinbarer Markdown-Dialekte

Markdown (III)

- John Gruber zeigte sich störrisch, was den Einbau zusätzlicher Features betraf: es kommt daraufhin zur Entwicklung zahlreicher Erweiterungen vieler z.T. unvereinbarer Markdown-Dialekte
- Seit 2014: Initiative zur Schaffung eines gemeinsamen Standards (ohne J. Gruber's Beteiligung):
CommonMark (commonmark.org)

Markdown (III)

- John Gruber zeigte sich störrisch, was den Einbau zusätzlicher Features betraf: es kommt daraufhin zur Entwicklung zahlreicher Erweiterungen vieler z.T. unvereinbarer Markdown-Dialekte
- Seit 2014: Initiative zur Schaffung eines gemeinsamen Standards (ohne J. Gruber's Beteiligung):
CommonMark (commonmark.org)
- CommonMark Spezifikation ist in CommonMark geschrieben

Markdown (III)

- John Gruber zeigte sich störrisch, was den Einbau zusätzlicher Features betraf: es kommt daraufhin zur Entwicklung zahlreicher Erweiterungen vieler z.T. unvereinbarer Markdown-Dialekte
- Seit 2014: Initiative zur Schaffung eines gemeinsamen Standards (ohne J. Gruber's Beteiligung):
CommonMark (commonmark.org)
- CommonMark Spezifikation ist in CommonMark geschrieben
- Referenz-Implementierungen (nur Konvertierung nach HTML) in C und JavaScript liegen bereits vor

Markdown (IV)

Syntax: Zwölf bis fünfzehn Grund-Regeln erfüllen bereits 90% aller Bedürfnisse:

Markdown (IV)

Syntax: Zwölf bis fünfzehn Grund-Regeln erfüllen bereits 90% aller Bedürfnisse:

- Beispiel: *“Absätze sind durch Leerzeilen voneinander getrennt.”*

Markdown (IV)

Syntax: Zwölf bis fünfzehn Grund-Regeln erfüllen bereits 90% aller Bedürfnisse:

- Beispiel: “*Absätze sind durch Leerzeilen voneinander getrennt.*”
- Beispiel: *kursiv*, **fett**, ***fettkursiv***, `inlinecode` geht so:

```
*kursiv*, **fett**, ***fettkursiv***, `inlinecode`
```

Markdown (IV)

Syntax: Zwölf bis fünfzehn Grund-Regeln erfüllen bereits 90% aller Bedürfnisse:

- Beispiel: *“Absätze sind durch Leerzeilen voneinander getrennt.”*
- Beispiel: *kursiv*, **fett**, ***fettkursiv***, `inlinecode` geht so:

```
*kursiv*, **fett**, ***fettkursiv***, `inlinecode`
```

- Beispiel: ein Link geht so:

```
[ein Link] (http://froscon.de/)
```

Markdown (V)

Syntax, Grund-Regeln, Fortsetzung:

Markdown (V)

Syntax, Grund-Regeln, Fortsetzung:

- Beispiel: Bild  geht so:

```
![Bild-Txt] (https://www.froscon.de/fileadmin/template/_src/FroSCon/FrontEnd/FroSCon/img/froscon_logo.png "Alternativ-Text")
```

(“Bild-Txt” ist optional. Eckige Klammern können auch leer bleiben. “Alternativ-Text” kann komplett entfallen.)

Markdown (V)

Syntax, Grund-Regeln, Fortsetzung:

- Beispiel: Bild  geht so:

```
![Bild-Txt](https://www.froscon.de/fileadmin/template/_src/FroSCon/FrontEnd/FroSCon/img/froscon_logo.png "Alternativ-Text")
```

(“Bild-Txt” ist optional. Eckige Klammern können auch leer bleiben. “Alternativ-Text” kann komplett entfallen.)

- Beispiel: Überschriften erster, zweiter, sechster Ordnung – Einleiten durch entsprechende Anzahl von #-Zeichen:

```
# Überschrift erster Ordnung  
[...]  
##### Überschrift sechster Ordnung
```

Markdown (VI)

Syntax, Grund-Regeln, Fortsetzung:

- Beispiel: Tabellen sind praktisch “ASCII-Art” (hier Methode `simple_table`):

```
+-----+-----+-----+
| Sorte   | Preis   | Vorrat  |
+=====+=====+=====+
| Bananen | 2,97 €  | nein    |
+-----+-----+-----+
| Erdbeeren | 1,49 €  | ja      |
+-----+-----+-----+
| Himbeeren | 3,11 €  | ja      |
+-----+-----+-----+
```


Markdown (VII)

- gerenderte Tabelle:

Sorte	Preis	Vorrat
Bananen	2,97 €	nein
Erdbeeren	1,49 €	ja
Himbeeren	3,11 €	ja

Markdown (VIII)

- Beispiel: Listen sind ebenfalls ganz leicht. Verschachteln von Listen durch Einrücken um jeweils 4 Leerzeichen pro zusätzlicher Ebene:

```
1. eins
1. zwei
  i. römisch 1
  i. römisch 2
    * etwas
    * was anderes
  i. römisch 3
1. drei
```

Markdown (IX)

Verschachtelte Liste, gerendert:

- ① eins
- ② zwei
 - ① römisch 1
 - ② römisch 2
 - etwas
 - was anderes
 - ③ römisch 3
- ③ drei

Pandoc

“Schweizer Offiziersmesser zur Konvertierung von Dokumenten-Formaten” (I)

“Schweizer Offiziersmesser zur Konvertierung von Dokumenten-Formaten” (I)



Figure 1: Pandoc 2006: Sehr nützliches kleines Tool!

“Schweizer Offiziersmesser zur Konvertierung von Dokumenten-Formaten” (II)

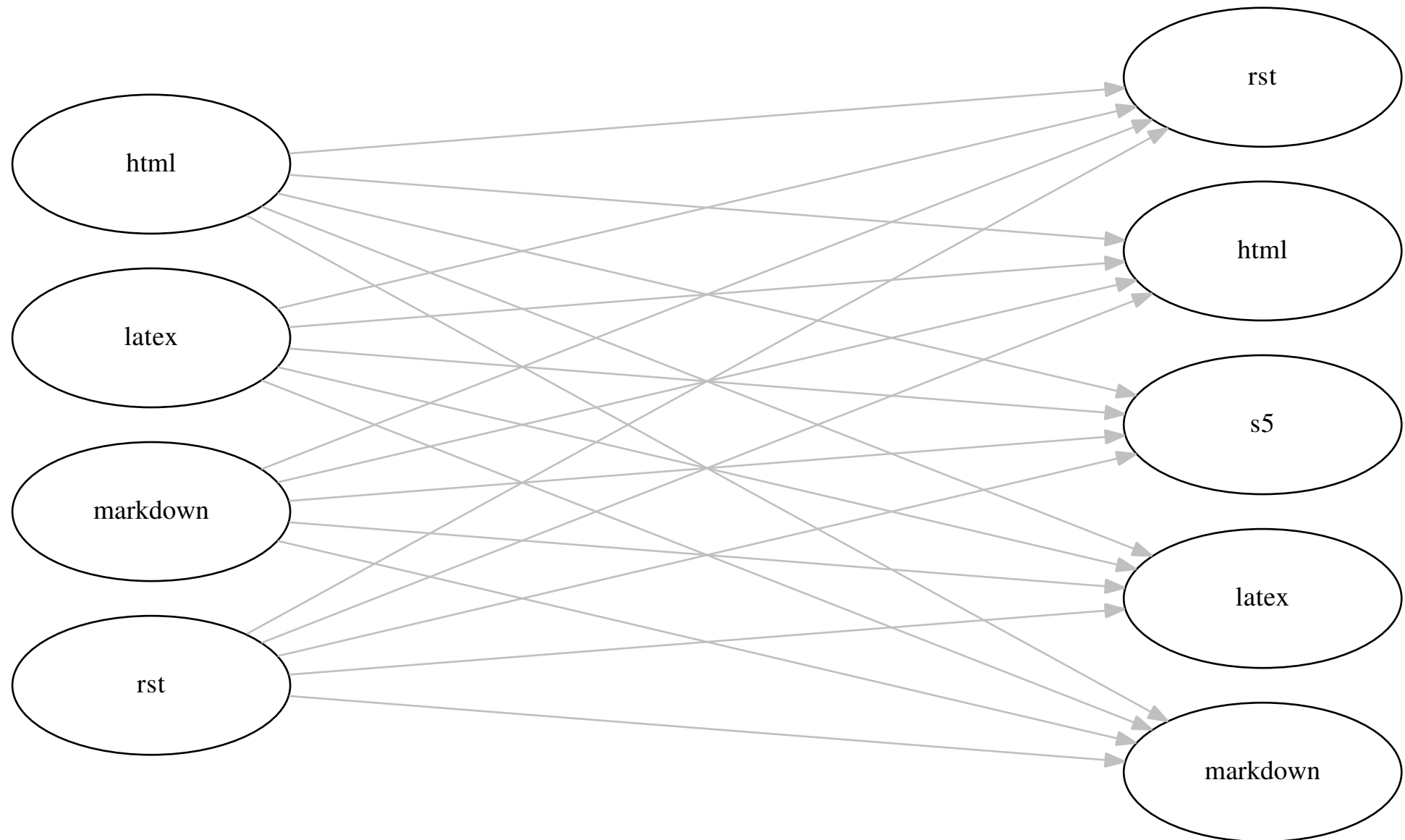
“Schweizer Offiziersmesser zur Konvertierung von Dokumenten-Formaten” (II)



Figure 2: Pandoc 2015: Unhandlich?

Darstellung ohne Vergleich aus Rüstungs-Industrie (I)

Darstellung ohne Vergleich aus Rüstungs-Industrie (I)



Darstellung ohne Vergleich aus Rüstungs-Industrie (II)

Darstellung ohne Vergleich aus Rüstungs-Industrie (II)

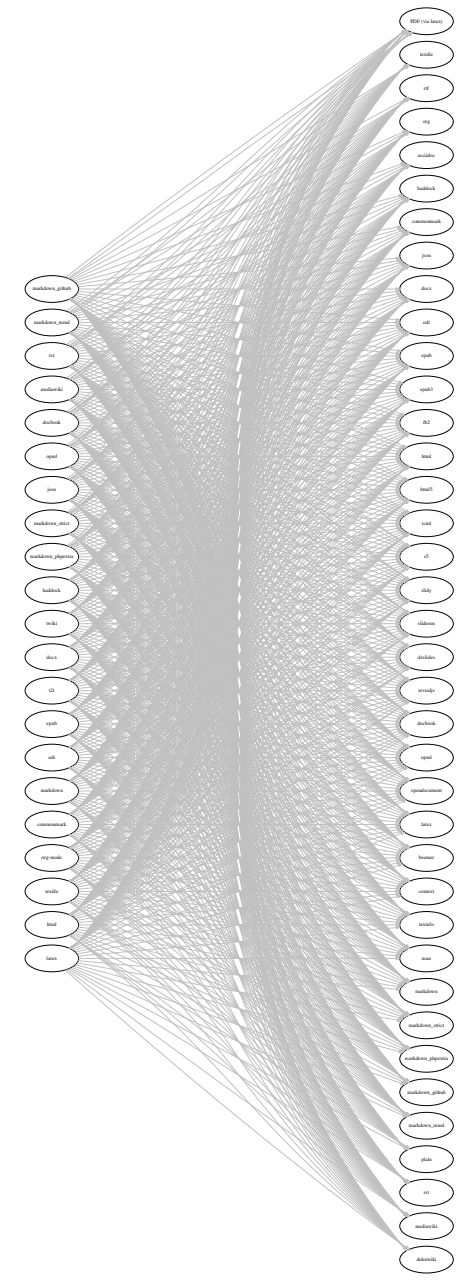


Figure 4. Pandoc 2015: Unhandlich?

Geschichtliches zu Pandoc

Geschichtliches zu Pandoc

Pandoc ist geschrieben von einem Berkeley-Professor für Philosophie und Logik

Geschichtliches zu Pandoc

Pandoc ist geschrieben von einem Berkeley-Professor für Philosophie und Logik
John MacFarlane startete das Projekt 2006 ursprünglich, um Haskell zu lernen

Geschichtliches zu Pandoc

Pandoc ist geschrieben von einem Berkeley-Professor für Philosophie und Logik
John MacFarlane startete das Projekt 2006 ursprünglich, um Haskell zu lernen
Vor allem in den letzten beiden Jahren: Zulauf neuer Entwickler, die weitere Input- und Output-Format einbrachten

Pandoc: Input-Formate (I)

Pandoc kann als *INPUT* nicht nur Markdown (alle Dialekte, inkl. CommonMark) verarbeiten, sondern auch:

- DocBook

Pandoc: Input-Formate (I)

Pandoc kann als *INPUT* nicht nur Markdown (alle Dialekte, inkl. CommonMark) verarbeiten, sondern auch:

- DocBook
- EPUB

Pandoc: Input-Formate (I)

Pandoc kann als *INPUT* nicht nur Markdown (alle Dialekte, inkl. CommonMark) verarbeiten, sondern auch:

- DocBook
- EPUB
- HTML (mit Einschränkungen)

Pandoc: Input-Formate (I)

Pandoc kann als *INPUT* nicht nur Markdown (alle Dialekte, inkl. CommonMark) verarbeiten, sondern auch:

- DocBook
- EPUB
- HTML (mit Einschränkungen)
- JSON

Pandoc: Input-Formate (I)

Pandoc kann als *INPUT* nicht nur Markdown (alle Dialekte, inkl. CommonMark) verarbeiten, sondern auch:

- DocBook
- EPUB
- HTML (mit Einschränkungen)
- JSON
- L^AT_EX (mit Einschränkungen)

Pandoc: Input-Formate (II)

weiter: mögliche Pandoc-Eingabe-Formate

- MediaWiki

Pandoc: Input-Formate (II)

weiter: mögliche Pandoc-Eingabe-Formate

- MediaWiki
- orgmode

Pandoc: Input-Formate (II)

weiter: mögliche Pandoc-Eingabe-Formate

- MediaWiki
- orgmode
- reStructuredText

Pandoc: Input-Formate (II)

weiter: mögliche Pandoc-Eingabe-Formate

- MediaWiki
- orgmode
- reStructuredText
- tags2text

Pandoc: Input-Formate (II)

weiter: mögliche Pandoc-Eingabe-Formate

- MediaWiki
- orgmode
- reStructuredText
- tags2text
- TWiki

Pandoc: Input-Formate (II)

weiter: mögliche Pandoc-Eingabe-Formate

- MediaWiki
- orgmode
- reStructuredText
- tags2text
- TWiki
- textile

Pandoc: Input-Formate (II)

weiter: mögliche Pandoc-Eingabe-Formate

- MediaWiki
- orgmode
- reStructuredText
- tags2text
- TWiki
- textile
- DOCX (!!!)

Pandoc: Ausgabe-Formate

Ausgabe-Formate von Pandoc außer als Input-Formate gelistete – nämlich:

Markdown-Dialekte, DocBook, EPUB/EPUB3, HTML, JSON, L^AT_EX, MediaWiki, orgmode, reStructuredText – auch:

- ASCII doc

Pandoc: Ausgabe-Formate

Ausgabe-Formate von Pandoc außer als Input-Formate gelistete – nämlich:
Markdown-Dialekte, DocBook, EPUB/EPUB3, HTML, JSON, L^AT_EX, MediaWiki, orgmode, reStructuredText – auch:

- ASCII doc
- PDF-Folien: Beamer, wie diese (?) Präsentation

Pandoc: Ausgabe-Formate

Ausgabe-Formate von Pandoc außer als Input-Formate gelistete – nämlich:
Markdown-Dialekte, DocBook, EPUB/EPUB3, HTML, JSON, L^AT_EX, MediaWiki, orgmode, reStructuredText – auch:

- ASCII doc
- PDF-Folien: Beamer, wie diese (?) Präsentation
- HTML-Folien: RevealJS, wie diese (?) Präsentation; S5; Slidy; Slideous; DZSlides

Pandoc: Ausgabe-Formate

Ausgabe-Formate von Pandoc außer als Input-Formate gelistete – nämlich:
Markdown-Dialekte, DocBook, EPUB/EPUB3, HTML, JSON, L^AT_EX, MediaWiki, orgmode, reStructuredText – auch:

- ASCII doc
- PDF-Folien: Beamer, wie diese (?) Präsentation
- HTML-Folien: RevealJS, wie diese (?) Präsentation; S5; Slidy; Slideous; DZSlides
- ODT (LibreOffice/OpenOffice)

Pandoc: Ausgabe-Formate

Ausgabe-Formate von Pandoc außer als Input-Formate gelistete – nämlich:
Markdown-Dialekte, DocBook, EPUB/EPUB3, HTML, JSON, L^AT_EX, MediaWiki, orgmode, reStructuredText – auch:

- ASCII doc
- PDF-Folien: Beamer, wie diese (?) Präsentation
- HTML-Folien: RevealJS, wie diese (?) Präsentation; S5; Slidy; Slideous; DZSlides
- ODT (LibreOffice/OpenOffice)
- DOCX

Pandoc: Ausgabe-Formate

Ausgabe-Formate von Pandoc außer als Input-Formate gelistete – nämlich:
Markdown-Dialekte, DocBook, EPUB/EPUB3, HTML, JSON, L^AT_EX, MediaWiki, orgmode, reStructuredText – auch:

- ASCII doc
- PDF-Folien: Beamer, wie diese (?) Präsentation
- HTML-Folien: RevealJS, wie diese (?) Präsentation; S5; Slidy; Slideous; DZSlides
- ODT (LibreOffice/OpenOffice)
- DOCX
- DokuWiki

Pandoc: Ausgabe-Formate

Ausgabe-Formate von Pandoc außer als Input-Formate gelistete – nämlich:
Markdown-Dialekte, DocBook, EPUB/EPUB3, HTML, JSON, L^AT_EX, MediaWiki, orgmode, reStructuredText – auch:

- ASCII doc
- PDF-Folien: Beamer, wie diese (?) Präsentation
- HTML-Folien: RevealJS, wie diese (?) Präsentation; S5; Slidy; Slideous; DZSlides
- ODT (LibreOffice/OpenOffice)
- DOCX
- DokuWiki
- ICML (InDesign!!)

Pandoc: Ausgabe-Formate

Ausgabe-Formate von Pandoc außer als Input-Formate gelistete – nämlich:
Markdown-Dialekte, DocBook, EPUB/EPUB3, HTML, JSON, L^AT_EX, MediaWiki, orgmode, reStructuredText – auch:

- ASCII doc
- PDF-Folien: Beamer, wie diese (?) Präsentation
- HTML-Folien: RevealJS, wie diese (?) Präsentation; S5; Slidy; Slideous; DZSlides
- ODT (LibreOffice/OpenOffice)
- DOCX
- DokuWiki
- ICML (InDesign!!)
- RTF

Pandoc: Ausgabe-Formate

Ausgabe-Formate von Pandoc außer als Input-Formate gelistete – nämlich:
Markdown-Dialekte, DocBook, EPUB/EPUB3, HTML, JSON, L^AT_EX, MediaWiki, orgmode, reStructuredText – auch:

- ASCII doc
- PDF-Folien: Beamer, wie diese (?) Präsentation
- HTML-Folien: RevealJS, wie diese (?) Präsentation; S5; Slidy; Slideous; DZSlides
- ODT (LibreOffice/OpenOffice)
- DOCX
- DokuWiki
- ICML (InDesign!!)
- RTF
- texinfo

Pandoc: Ausgabe-Formate

Ausgabe-Formate von Pandoc außer als Input-Formate gelistete – nämlich:
Markdown-Dialekte, DocBook, EPUB/EPUB3, HTML, JSON, L^AT_EX, MediaWiki, orgmode, reStructuredText – auch:

- ASCII doc
- PDF-Folien: Beamer, wie diese (?) Präsentation
- HTML-Folien: RevealJS, wie diese (?) Präsentation; S5; Slidy; Slideous; DZSlides
- ODT (LibreOffice/OpenOffice)
- DOCX
- DokuWiki
- ICML (InDesign!!)
- RTF
- texinfo
- manpages (GROFF)

Zugaben!

Zugabe: Mathe- und Formel-Satz

- Per Kommandozeilen-Zusatz. Mehrere Möglichkeiten:
- `--gladtex`, `--jsmath`, `--katex`, `--latexmathml`, `--mathjax`, `--mathml`, `--mimetex` oder `--webtex`.
(falls `--mathml`: funktioniert nicht in Chrome, nur in Firefox)

$$\frac{2+dc}{9}$$

$$\sum_{i=1}^n i = \frac{n(n+1)}{2}$$

$$\prod_{i=1}^{n+1} i = 1 \cdot 2 \cdot \dots \cdot n \cdot (n+1)$$

$$\int_0^3 x^2 dx = 9$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} = 0$$

Zugabe: Zitate und Bibliografien

Zugabe: Zitate und Bibliografien

Man braucht drei verschiedene Zutaten, um Literatur-Verzeichnisse zu erzeugen:

- 1 Kurz-Verweise (in einem spezifischen Format) auf Literatur im Markdown-Text.

Zugabe: Zitate und Bibliografien

Man braucht drei verschiedene Zutaten, um Literatur-Verzeichnisse zu erzeugen:

- 1 Kurz-Verweise (in einem spezifischen Format) auf Literatur im Markdown-Text.
- 2 Eine separate Datei (in einem spezifischen Format), welche zu den Kurz-Verweisen die Detail-Infos enthält.

Zugabe: Zitate und Bibliografien

Man braucht drei verschiedene Zutaten, um Literatur-Verzeichnisse zu erzeugen:

- 1 Kurz-Verweise (in einem spezifischen Format) auf Literatur im Markdown-Text.
- 2 Eine separate Datei (in einem spezifischen Format), welche zu den Kurz-Verweisen die Detail-Infos enthält.
- 3 Eine **CSL**-Datei (*Citation Style Language*), welche die Definitionen für die Formatierung des Literatur-Verzeichnisses enthält.

Zugabe: Zitate und Bibliografien

Man braucht drei verschiedene Zutaten, um Literatur-Verzeichnisse zu erzeugen:

- 1 Kurz-Verweise (in einem spezifischen Format) auf Literatur im Markdown-Text.
- 2 Eine separate Datei (in einem spezifischen Format), welche zu den Kurz-Verweisen die Detail-Infos enthält.
- 3 Eine **CSL**-Datei (*Citation Style Language*), welche die Definitionen für die Formatierung des Literatur-Verzeichnisses enthält.

Zugabe: Zitate und Bibliografien

Man braucht drei verschiedene Zutaten, um Literatur-Verzeichnisse zu erzeugen:

- 1 Kurz-Verweise (in einem spezifischen Format) auf Literatur im Markdown-Text.
- 2 Eine separate Datei (in einem spezifischen Format), welche zu den Kurz-Verweisen die Detail-Infos enthält.
- 3 Eine **CSL**-Datei (*Citation Style Language*), welche die Definitionen für die Formatierung des Literatur-Verzeichnisses enthält.

Diese drei Zutaten muss man mit entsprechenden Parametern beim Pandoc-Aufruf referenzieren:

```
--bibliography=/path/to/my.bib    \  
--csl=/path/to/my.csl           \  
--filter=/path/to/pandoc-citeproc \  

```

Zugabe: Zitate und Bibliografien

Man braucht drei verschiedene Zutaten, um Literatur-Verzeichnisse zu erzeugen:

- ➊ Kurz-Verweise (in einem spezifischen Format) auf Literatur im Markdown-Text.
- ➋ Eine separate Datei (in einem spezifischen Format), welche zu den Kurz-Verweisen die Detail-Infos enthält.
- ➌ Eine **CSL**-Datei (*Citation Style Language*), welche die Definitionen für die Formatierung des Literatur-Verzeichnisses enthält.

Diese drei Zutaten muss man mit entsprechenden Parametern beim Pandoc-Aufruf referenzieren:

```
--bibliography=/path/to/my.bib    \  
--csl=/path/to/my.csl            \  
--filter=/path/to/pandoc-citeproc \  

```

- Siehe mein GitHub-Repository: **Pandoc-Tests mit unterschiedlichen CSL-Dateien**