DHPa

gitlab

Semi-struktur

arser

b.lrz.de/badw-it/DHPars

rierte Daten

Formale Gr
zur Retro-D
bibliograph

rser - dhparser.readth

DH

rammatiken
Digitalisierung
nischer Daten

hedocs.io

lParser

g

- Daten in **schemat**
- aber **von Mensch**
- daher viele **Ausna**
- in Textformaten ni
  ohne semantisch
- meist nur ein unvo
  kein explizites R

Beispiele:

(Fach-)Bibliografien
Druck-Wörterbüche

tischer Form
nenhand gepflegt
ahmen, Sonderfälle
niedergelegt,
he Auszeichnungen
vollständiges oder gar
Regelwerk vorhanden

n in Buchform,
ner, Kataloge

Parser

dhparser.readthedocs.i

(Baum-)Tra

dhparser.readthedocs.io/

Testu

dhparser.readthedocs.io/

[1] für Nicht-Baumstrukturen, z.B Graphen oder, wi
Bearbeitungsschritte außerhalb von DHParser ang

Beispiel 1: I

Albert Weber: Bibliograpl
aus dem östlichen Europa

```
Titel: SOZIALDEMOKRA
IN BULGARIEN
Erscheinungsort: So
Erscheinungsbeginn:
Ethnische Gemeinscha
```

ergenerator

cs.io/en/latest/manuals/01_ebnf.html

ansformationen [1)]

s.io/en/latest/manuals/03_transform.html

umgebung

.io/en/latest/manuals/05_testing.html

wie in den Beispielen unten LOD, müssen noch weitere angefügt werden.

Der günstige Fall:

phie deutschsprachiger Periodika
ba, IOS Regensburg 2013.

RATISCHE ARBEITERPARTEI

ofia
: 1907, 1910[AW4]
haft: Bulgariendeutsche

a

Formale Gram

- taugen wie regulä
  Erkennung von T
- **leistungsfähiger**
  und übersichtliche

Herausforderunger

- **Steile Lernkurve**
- ohne explizites Re

nmatiken

äre Ausdrücke zur
Textmustern
r (durch Rekursion)
ner als reg. Ausdrücke

en:

re für Nicht-Informatiker
egelwerk (siehe

Ethnische Gemeinscha
GLP ID: 0016

- **Daten vom Autor vorl**
  (beugt auch Fehlinterpre
- Klare und weitgehend **e**
- Mäßiger **Variantenreic**
  durch die teils komplizie
  verzeichneten Werke

→ **Arbeitsaufwand Par**

er

Beispiel 2: D

Ralf Schönberger et al.: R
Mittelalters aus dem Ber
angrenzender Gebiete

B1700-150/65

haft: Bulgariendeutsche

rbereitet für die Digitalisierung
retationen der Techniker vor!)
eindeutige Struktur
ichtum, sachlich bedingt u.a.
erte Erscheinungsgeschichte der

rserbau: 2-3 Werktage

Der ungünstige Fall:

Repertorium edierter Texte des
reich der Philosophie und

g

r

oben) muss die Gr
Versuch-und-Irr
(→ Testumgebung

- **keine Erfolgsgar**
  überhaupt genug S
  um ihre Syntax fo
  nicht von Anfang

Dr. A

Ingo

rammatik in einem
rtumsverfahren
ng) rekonstruiert werden
rantie: Ob die Daten
Struktur enthalten,
ormal zu beschreiben, ist
g an absehbar

Dr. Eckhart Arnold, Bayerische A
www.e
Albert Weber, Leibnitz-Institut für C
leibniz-ios.de/pers
o Frank M.A., Leibnitz-Institut für O

```
    B1700-150/65
ED: Chantelou, Claudius.
IN: Bibliotheca Patrum a
Patrum de christiana ac
opuscula, I, Paris 1662.
AT: Panereticon, pars pr
de sanctis complectens
Bernardus Claraevallensi
⇨ B1700-170/20
```

n

st

- Daten nicht präpariert:
- Teils **uneindeutige Felda**
- **Idiosynkratischer Varia**
  Literatur- Orts- und Zeita
- **Fehler**, z.B. Rechtschreibf

→ machbar, aber **Arbeitsa**

Akademie der Wissenschaften, Münc
.eckhartarnold.de
Ost- und Südosteuropa Forschung, R
ersonen/details/albert-weber
Ost- und Südosteuropa Forschung, R

s.
ascetica sive selecta veterum
c religiosa perfectione
2.
orima, Sermones de tempore et

sis: Vita s. Malachiae.

t: Sinn **machmal nur zu errater**
dabgrenzungen
iantenreichtum (insbesondere bei
tangaben)
bfehler in Schlüsselwörtern!
saufwand schwer abschätzbar

nchen

Regensburg

Regensburg

en.

Ingo

Ost- und Südosteuropa Forschung, Re

personen/details/ingo-frank

Regensburg