

Digitalisierung von Büchern

Rückerts Werke in 12 Bänden

ein Erfahrungsbericht

Who? Peter Koppatz

From? Supported Distance Learning – Besser lehren
Beratung, Konzeption, Schulung

When? Linuxtag in Chemnitz am 16. März, 2013



Projekt: rueckert.gesammelte-werke.org

- 1 Idee und Motivation für das Projekt
- 2 Stand der Dinge: ein Blick in andere Projekte
- 3 Stärken und Schwächen von Digitalisierungsprojekten
- 4 Ziele des Rückert-Projektes
- 5 Die Umsetzung
- 6 Was ist noch zu tun?
- 7 Diskussion und ggf. Demonstration einiger Arbeitsschritte



Lesen bildet. . .

© Randy Glasbergen
www.glasbergen.com



Das nennt man Lesen. So installieren
Menschen neue Software in ihrem Gehirn.

PK

Idee und Motivation für das Projekt

Eine Radiosendung und ihre Folgen:

- 1 www.dradio.de/dlf/sendungen/dasfeature/1705152/
- 2 Ein Gang ins Antiquariat . . . Treffer!
- 3 Eine zwölfbändige Ausgabe 😊 mit nur elf Bänden ☹️ geht in meinen Besitz über.
- 4 Experimentieren mit neuer Software
- 5 Ein Trick, die Bücher nicht nur ins Regal zu stellen, sondern auch zu lesen.
- 6 Das Angebot kostenloser Literatur erweitern.
- 7 Eine qualitativ hochwertige Ausgabe, die Lust aufs Stöbern macht.

Ich bin schon da ...

- Hase & Igel
oder
- Es gibt doch schon andere Anbieter!

Es gibt doch schon andere Anbieter!

Als da wären . . .

- GoogleBooks



Es gibt doch schon andere Anbieter!

Als da wären . . .

- GoogleBooks
- Gutenberg [org]



Es gibt doch schon andere Anbieter!

Als da wären . . .

- GoogleBooks
- Gutenberg [org]
- Gutenberg [de]

Es gibt doch schon andere Anbieter!

Als da wären . . .

- GoogleBooks
- Gutenberg [org]
- Gutenberg [de]
- OpenLibrary

Es gibt doch schon andere Anbieter!

Als da wären ...

- GoogleBooks
- Gutenberg [org]
- Gutenberg [de]
- OpenLibrary
- OpenArchive

Das aktuelle Angebot (2013)

- Angebotene Titel insgesamt und in deutscher Sprache
- Anteile verschiedener Genres deutschsprachiger Literatur auf gutenberg.org
- Wo findet man Friedrich Rückert bei anderen Anbietern?



Recherche vor dem Kauf!

Angeborene Titel insgesamt und in deutscher Sprache

Anbieter	Anzahl	
	gesamt	deutschsprachig
gutenberg.org	42.000	731
projekt.gutenberg.de	6746	6746
archive.org	3.907.235	227
openlibrary.org	1.000.000	12.665

Rubriken

Anteile verschiedener Genres deutschsprachiger Literatur auf
gutenberg.org

Rubrik	Anzahl
Drama	23
Märchen	69
Lyrik	39
Prosa	365
Sachbücher	235
Summe	731

Friedrich Rückert

Wo ist er zu finden?

Rubrik	Anzahl
gutenberg.org	1
projekt.gutenberg.de	5
archive.org	57
openlibrary.org	43

Doppelte Einträge und suchmaschinenabhängige Ergebnisse sind möglich.

Was geht – was geht nicht?

- Am Beispiel von Kafkas „Die Verwandlung“
- Nutzungsbedingungen
- Nur als eBook und doch nur halboffen?

Ausgabeformate/Export

Am Beispiel von Kafkas „Die Verwandlung“

gutenberg.org (1), projekt.gutenberg.de (2),
archive.org (3), openlibrary.org (4)

Format	1	2	3	4
online	x	x	x	x
html	x	-	-	-
Text	x	-	-	x
PDF	-	-	-	x
epub	x	-	x	x
kindle	x	-	-	x
Plucker	x	-	x	-
QiOO Mobil	x	-	-	-
DAISY	-	-	-	x
MOBI	-	-	-	x
DjVu	-	-	-	x

Nutzungsbedingungen: 

Warum die Frage, wenn ich es einmalig als PDF speichern will?

Rückert, Friedrich/Rückert, Heinrich:
Gesammelte poetische Werke in zwölf Bänden
Frankfurt a. M. (1868)

Erstes Bild

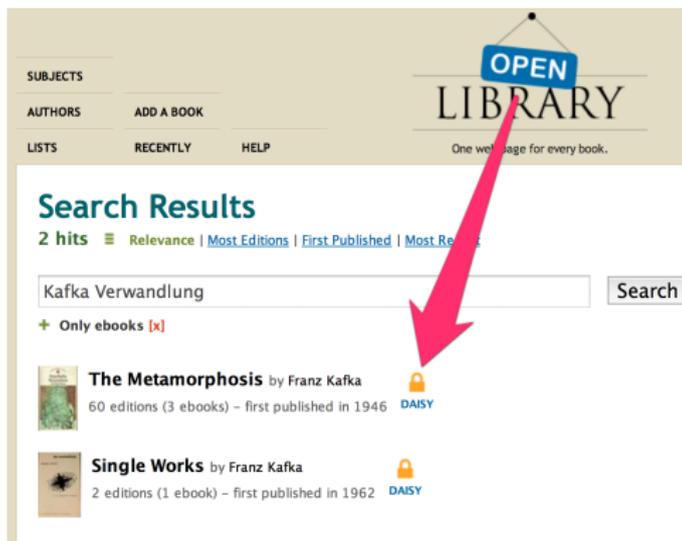
Letztes Bild

Bitte bestätigen Sie die nicht kommerzielle Nutzung der Daten



Quelltext herunterladen

Nur als e-Book und doch nur halboffen? 😞



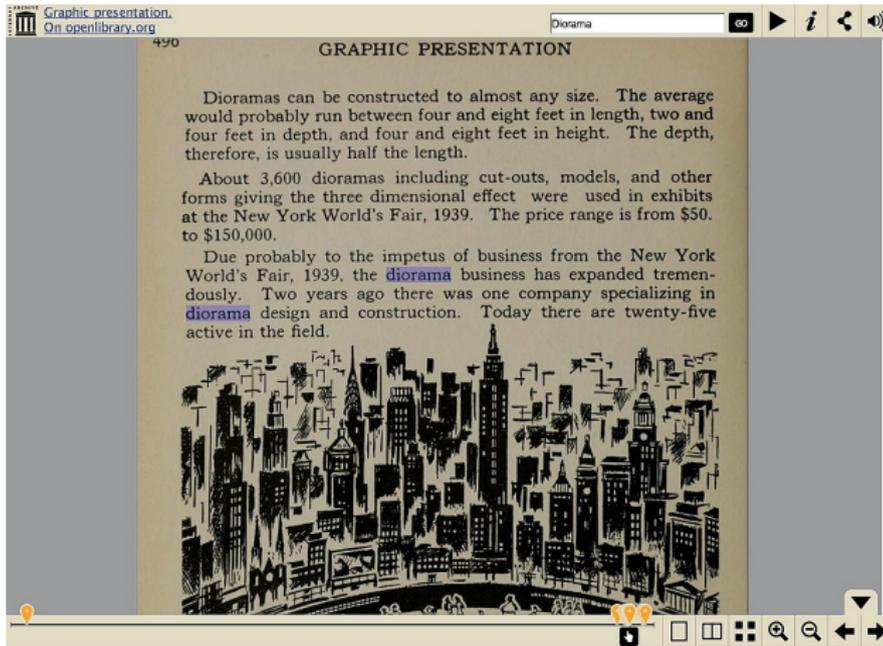
The screenshot shows the Open Library website interface. At the top, there is a navigation bar with links for SUBJECTS, AUTHORS, ADD A BOOK, LISTS, RECENTLY, and HELP. The Open Library logo is prominently displayed, featuring a blue sign that says 'OPEN' hanging from a string above the word 'LIBRARY'. Below the logo is the tagline 'One webpage for every book.'.

The main content area is titled 'Search Results' and shows '2 hits'. There are sorting options: Relevance, Most Editions, First Published, and Most Recent. A search bar contains the text 'Kafka Verwandlung' and a 'Search' button. Below the search bar, there is a filter option '+ Only ebooks [x]'. Two search results are listed:

- The Metamorphosis** by Franz Kafka. It has 60 editions (3 ebooks) and was first published in 1946. A red padlock icon indicates it is not fully open access, and the text 'DAISY' is shown next to it. A large red arrow points from the search bar to this result.
- Single Works** by Franz Kafka. It has 2 editions (1 ebook) and was first published in 1962. A red padlock icon indicates it is not fully open access, and the text 'DAISY' is shown next to it.



- auf archive.org
- auf projekt.gutenberg.de
- auf gutenberg.org



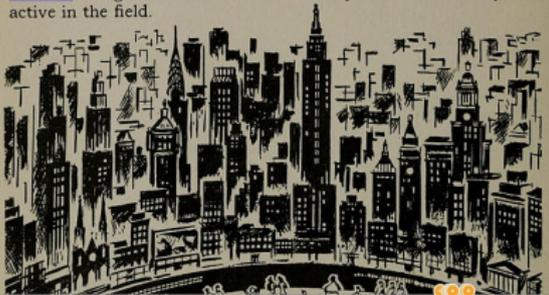
490

GRAPHIC PRESENTATION

Dioramas can be constructed to almost any size. The average would probably run between four and eight feet in length, two and four feet in depth, and four and eight feet in height. The depth, therefore, is usually half the length.

About 3,600 dioramas including cut-outs, models, and other forms giving the three dimensional effect were used in exhibits at the New York World's Fair, 1939. The price range is from \$50. to \$150,000.

Due probably to the impetus of business from the New York World's Fair, 1939, the **diorama** business has expanded tremendously. Two years ago there was one company specializing in **diorama** design and construction. Today there are twenty-five active in the field.



PK

Online lesen

in kleinem Fenster? 😞 (projekt.gutenberg.de)

Franz Kafka: Die Verwandlung - Kapitel 3

Navigation:

Kein flexibles Layout!
Abgeschnittener Text,
wenn Bildschirm zu klein.



Dann aber sagte er sich: »Ehe es einviertel acht schlägt, muß ich unbedingt das Bett voll bis dahin jemand aus dem Geschäft kommen, um nach mir zu fragen, denn das Geschäft sich nun daran, den Körper in seiner ganzen Länge vollständig gleichmäßig aus dem Bett Weise aus dem Bett fallen ließ, blieb der Kopf, den er beim Fall scharf heben wollte, voraus sein; dem würde wohl bei dem Fall auf den Teppich nichts geschehen. Das größte Bedr Krach, den es geben müßte und der wahrscheinlich hinter allen Türen wenn nicht Schreck mußte aber gewagt werden.

Als Gregor schon zur Hälfte aus dem Bette ragte – die neue Methode war mehr ein Spie ruckweise zu schaukeln – , fiel ihm ein, wie einfach alles wäre, wenn man ihm zu Hilfe kä Vater und das Dienstmädchen – hätten vollständig genügt; sie hätten ihre Arme nur unter dem Bett schälen, sich mit der Last niederbeugen und dann bloß vorsichtig dulden müssen vollzog, wo dann die Beinchen hoffentlich einen Sinn bekommen würden. Nun, ganz abge hätte er wirklich um Hilfe rufen sollen? Trotz aller Not konnte er bei diesem Gedanken ein L:

Schon war er so weit, daß er bei stärkerem Schaukeln kaum das Gleichgewicht noch erh entscheiden, denn es war in fünf Minuten einviertel acht, – als es an der Wohnungstür läute



Online lesen

in kleinem Fenster?  (gutenberg.org)

Online Reader

Die Verwandlung by Franz Kafka

[First Page](#) — [Previous Page](#) — Goto page:

— [Next Page](#) — [My Bookmarks](#) — [Download Book](#)

Er glitt wieder in seine frühere Lage zurück. »Dies frühzeitig aufstehen«, dachte er, »macht einen ganz blödsinnig. Der Mensch seinen Schlaf haben. Andere Reisende leben wie Haremsfrauen. V zum Beispiel im Laufe des Vormittags ins Gasthaus zurückgehe, erlangten Aufträge zu überschreiben, sitzen diese Herren erst Frühstück. Das sollte ich bei meinem Chef versuchen; ich würde Stelle hinausfliegen. Wer weiß übrigens, ob das nicht sehr gut wäre. Wenn ich mich nicht wegen meiner Eltern zurückhielte, ich längst gekündigt, ich wäre vor den Chef hingetreten und hätte Meinung von Grund des Herzens aus gesagt. Vom Pult hätte er fe müssen! Es ist auch eine sonderbare Art, sich auf das Pult zu von der Höhe herab mit dem Angestellten zu reden, der überdies Schwerhörigkeit des Chefs ganz nahe herantreten muß. Nun, die ist noch nicht gänzlich aufgegeben, habe ich einmal das Geld k um die Schuld der Eltern an ihn abzuzahlen -- es dürfte noch f sechs Jahre dauern --, mache ich die Sache unbedingt. Dann wird große Schnitt gemacht. Vorläufig allerdings muß ich aufstehen, Zug fährt um fünf.«



- Beispiel: Bayerische Staatsbibliothek
- Beispiel: archive.org

Volltextsuche in ...

Beispiel: Bayerische StaatsBibliothek



BSB Bayerische
StaatsBibliothek digital

Lyrische Gedichte. Buch 1, Vaterland.

Buch 2, Liebesfrühling

Autor / Hrsg.: Rückert, Friedrich ;
Rückert, Heinrich

Verlagsort: Frankfurt a. M. |

Erscheinungsjahr: (1868) | **Verlag:**
Sauerländer

Signatur: P.o.germ. 1184 em-1

Reihe: Gesammelte poetische Werke :
in zwölf Bänden ; 1

Permalink: <http://www.mdz-nbn-resolving.org/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb10311642-6>

Kein Volltext verfügbar.

[Suche im Band] [PDF-Download]
[OPAC]

Tex

Volltextsuche in ...

Volltextsuche: archive.org

Das er nimmer auf grünem Moos
Schatten kann streun uns beiden.
Wenn der Herbst die Lilien leicht,
Denk' und weine, so zu nicht
Ist un...
Wenn der Frühling aus Schnee und Eis

Moos GO

No matches were found.

PK

- Gute Volltextsuche
- Gute Lesbarkeit (online)
- Quellen im REST-Format
- Export in diverse Formate

Search Results

Search finished, found 2 page(s) matching the search query.

- **Der vermauerte Schatz.**

Der vermauerte Schatz. ===== 068 - 070
~~~~~ :: Im Wald ist voll be**moos**ter Zertrümmerter Stein' ein  
Platz; Dort stand das alte Kloster, Dort liegt der vermauerte Schatz.  
Auszog nach Morgenlanden Ein Ri...

- **Körner's Geist.**

Körner's Geist. ===== 77-78 ~~~~~ :: Bedeckt von  
**Moos** und Schorfe, Ein Eichbaum hoch und stark, Steht bei Wöbblin,  
dem Dorfe, In Mecklenburger Mark. Darunter ist von Steine Ein  
neues Grab gemacht, D'raus ste...

### Die Schreibfeder.

034 - 035

Wer hat Unterschleif getrieben,  
Mit der Feder wer geschrieben,  
Die zu meiner Schreiberei  
Eingeweihet war? Es verstauchte  
Mir die Feder, wer sie brauchte,  
Wer es auch gewesen sei.  
Hat die Frau darauf gesprochen:  
Hab' ich sie doch nicht zerbrochen,  
Und nur im Haushaltsbuch  
Hab ich hier sechs oder sieben  
Wicht'ge Zahlen angeschrieben,  
Wichtiger als mancher Spruch.  
„Freilich ist der Haushalt wichtig,  
Und die Poesie ist nichtig,  
Ich erkenn' es in Geduld.  
Aber kann ich mit ihr wieder  
Nie mehr schreiben Liebeslieder,  
Geb' ich dir allein die Schuld.“

Anmerkungen



# Ziele für gesammelte-werke.org

## Quellen im REST-Format

### Die Schreibfeder.

034 - 035

::

Wer hat Unterschleif getrieben,  
Mit der Feder wer geschrieben,  
Die zu meiner Schreiberei  
Eingeweiht war? Es verstauchte  
Mir die Feder, wer sie brauchte,  
Wer es auch gewesen sei.  
Hat die Frau darauf gesprochen:  
Hab' ich sie doch nicht zerbrochen,  
Und nur im Haushaltsbuch  
Hab ich hier sechs oder sieben  
Wicht'ge Zahlen angeschrieben,  
Wichtiger als mancher Spruch.  
„Freilich ist der Haushalt wichtig,  
Und die Poesie ist nichtig,  
Ich erkenn' es in Geduld.  
Aber kann ich mit ihr wieder  
Nie mehr schreiben Liebeslieder,  
Geb' ich dir allein die Schuld.“

### Anmerkungen

~

.. glossary::

Unterschleif  
veraltet: Unterschlagung

~siehe auch Wiktionary <<http://de.wiktionary.org/wiki/Unterschleif>>

# Ziele für gesammelte-werke.org

## Export in diverse Formate

```
(rueckert)tinship-2:buecher peterkoppatz$ make
Please use `make <target>' where <target> is one of
  html      to make standalone HTML files
  dirhtml   to make HTML files named index.html in directories
  singlehtml to make a single large HTML file
  pickle    to make pickle files
  json      to make JSON files
  htmlhelp  to make HTML files and a HTML help project
  qthelp    to make HTML files and a qthelp project
  devhelp   to make HTML files and a Devhelp project
  epub      to make an epub
  latex     to make LaTeX files, you can set PAPER=a4 or PAPER=letter
  latexpdf  to make LaTeX files and run them through pdflatex
  text      to make text files
  man       to make manual pages
  texinfo   to make Texinfo files
  info      to make Texinfo files and run them through makeinfo
  gettext   to make PO message catalogs
  changes   to make an overview of all changed/added/deprecated items
  linkcheck to check all external links for integrity
  doctest   to run all doctests embedded in the documentation (if enabled)
(rueckert)tinship-2:buecher peterkoppatz$ █
```

# Umsetzung . . .

- Scannen
- Texterkennung mit der Software: Tesseract
- Texterkennung in GoogleDocs
- Texterkennung auf eigenem Rechner
- Der Kommandozeilenaufruf
- Tesseract-Installation
- Tesseract für MacOS
- OCR für Fraktur
- Formatieren im REST-Format
- Transformation
- Datensicherung
- Kleine Helfer in Python
- Index
- Korrekturlesen



# Die Realisierung/Umsetzung

## Scannen

Bilder mit 300 DPI + Nachbearbeitung

- Ausschneiden der Gedichte, um Speicherplatz zu sparen
- Drehen, falls schräg eingescannt
- Dateiname vergeben: s001.jpg, s002.jpg, s003.jpg ...

# Texterkennung

mit der Software: Tesseract

„Wir sind die Guten . . .“

- Website: <http://code.google.com/p/tesseract-ocr/>



## tesseract-ocr

An OCR Engine that was developed at HP Labs between 1985 and 1995... and now at Google.

- Frakturschrift wird nicht direkt unterstützt, es gibt aber Erweiterungen.

# Texterkennung I von III

Texterkennung in GoogleDocs

## Ergebnis in Googelbooks (online)

*Kleines Flauenlob Frauen sind genannt vom Freuen Weil  
sich sreuen kann kein Mann Ohn ein Weib das stets vom  
neuen Seel und Leib ersreuen kann Wohlgesraut ist wohl-  
gesreuet Ungesreut ist ungesraut Wer der Frauen Auge  
scheuet Hat die Freude nie geschaut Wie ersreulich wo  
so sraulich Eine Frau geberdet sich So getreulich und so  
traulich Wie sich eine schmiegt an mich*

# Texterkennung II von III

Texterkennung auf eigenem Rechner

## Ergebnis Tesseract auf eigenem Rechner

*Kleines Franks-lob.*

*Frauen find genannt vom Neuen,  
Weil sich freuen kann sein Mann  
Ohn' ein Weib, dag stets von: neuen  
SeeP und Leib erfreuen sann.*

*Wohlgcfraut ist wohlgefueh  
Ungefreet ist ungefrau  
Wer der Frauen Auge schenkt,  
Hat die Freude nie geschaut.*

*Wie erfreulich, wo so fraulich  
Eine Frau geberdet sich,  
So getreulich und so traulich,  
Wie fich eine schntiegt an mich.*

# Texterkennung III von III

## Der Kommandozeilenaufruf

Aufruf mit:

```
tesseract frauenlob.png -l deu frauenlob
```



# Tesseract-Installation

<http://code.google.com/p/tesseract-ocr/wiki/Compiling>

## Abhängigkeiten prüfen, am Beispiel von Debian:

```
sudo apt-get install autoconf automake libtool  
sudo apt-get install libpng12-dev  
sudo apt-get install libjpeg62-dev  
sudo apt-get install libtiff4-dev  
sudo apt-get install zlib1g-dev
```

Zusätzlich muss Leptonica installiert sein.



# Tesseract-Installation

<http://code.google.com/p/tesseract-ocr/wiki/Compiling>

## Installations-Dreisatz

```
./autogen.sh  
./configure --prefix=$HOME/local/  
make install
```



# Tesseract für MacOS

<http://code.google.com/p/tesseract-ocr/wiki/Compiling>

Mit Unterstützung von «Homebrew»

```
install tesseract
```

```
ls /usr/local/Cellar/tesseract/<version>/share/tessdata
```

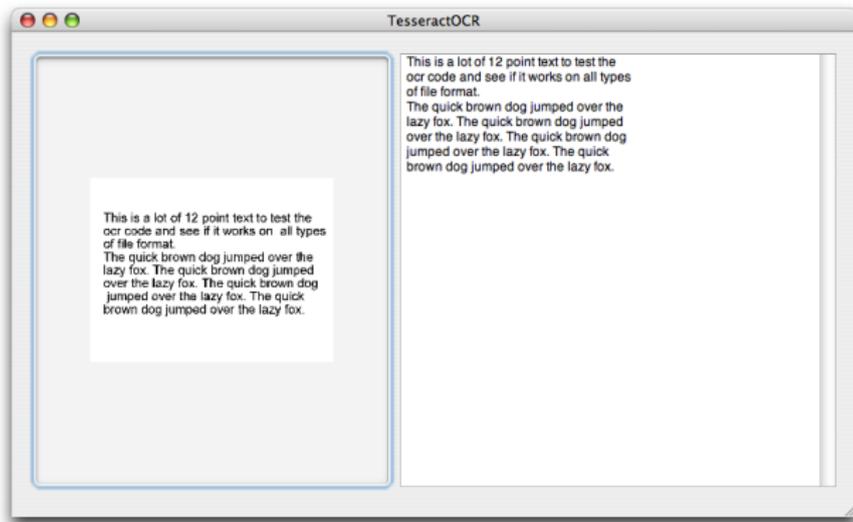
```
deu.traineddata
```

```
eng.traineddata
```



# Tesseract für MacOS

Ein GUI von <http://www.malcolmhardie.com/ocr/index.html>



# OCR für Fraktur

<http://code.google.com/p/tesseract-ocr/downloads/list>



[deu-frak.traineddata.gz](#)

German (Fraktur)  
language data for  
Tesseract (3.00 and  
up)

deu.traineddata kopieren

```
cp deu.traineddata <mein tesseract-Ordner>/tessdata
```

# Formatieren im REST-Format

<http://sphinx-doc.org/rest.html>

Die Schreibfeder.

034 - 035

::

Wer hat Unterschleif getrieben,  
Mit der Feder wer geschrieben,  
Die zu meiner Schreiberei  
Eingeweiht war? Es verstauchte  
Mir die Feder, wer sie brauchte,  
Wer es auch gewesen sei.  
Hat die Frau darauf gesprochen:  
Hab' ich sie doch nicht zerbrochen,  
Und nur im Haushaltsbuch  
Hab ich hier sechs oder sieben  
Wicht'ge Zahlen angeschrieben,  
Wichtiger als mancher Spruch.  
„Freilich ist der Haushalt wichtig,  
Und die Poesie ist nichtig,  
Ich erkenn' es in Geduld.  
Aber kann ich mit ihr wieder  
Nie mehr schreiben Liebeslieder,  
Geb' ich dir allein die Schuld.“

Anmerkungen

□

.. glossary::

Unterschleif  
veraltet: Unterschlagung

~siehe auch Wiktionary <<http://de.wiktionary.org/wiki/Unterschleif>>~

# Transformation

## Aufruf mit make

```
(rueckert)tinship-2:buecher peterkoppatz$ make html
sphinx-build -b html -d ../werke/doctrees source ../werke/html
Running Sphinx v1.1.3
loading pickled environment... done
building [html]: targets for 0 source files that are out of date
updating environment: 0 added, 0 changed, 0 removed
looking for now-outdated files... none found
no targets are out of date.

Build finished. The HTML pages are in ../werke/html.
(rueckert)tinship-2:buecher peterkoppatz$ █
```



# Datensicherung

## Mercurial/Rhodecode

RhodeCode

HOME » FLASK/RUECKERT » CHANGELOG - SHOWING 20 OUT OF 105 REVISIONS

Open new pull request

Show 20 revisions All branches

|                          |                        |                                  |                                                    |
|--------------------------|------------------------|----------------------------------|----------------------------------------------------|
| <input type="checkbox"/> | 104:94699777a721       | Bilder zu amrayllis und agnes    | 122                                                |
|                          | pkop (RhodeCode Admin) | Wed, 27 Feb 2013 13:57:39        | Parent 103:a0769f8b0271<br><a href="#">DETAILS</a> |
| <input type="checkbox"/> | 103:a0769f8b0271       | Amaryllis und Agnes              | 121                                                |
|                          | pkop (RhodeCode Admin) | Wed, 27 Feb 2013 06:53:34        | Parent 102:13a7c0eeef43<br><a href="#">DETAILS</a> |
| <input type="checkbox"/> | 102:13a7c0eeef43       | Amaryllis vollendet              | 40                                                 |
|                          | pkop (RhodeCode Admin) | Tue, 01 Jan 2013 19:58:13        | Parent 101:2cb2dffa346c<br><a href="#">DETAILS</a> |
| <input type="checkbox"/> | 101:2cb2dffa346c       | Neues vom 16. Dezember 2012      | 13                                                 |
|                          | pkop (RhodeCode Admin) | Sun, 16 Dec 2012 10:28:39        | Parent 100:77e076b34d80<br><a href="#">DETAILS</a> |
| <input type="checkbox"/> | 100:77e076b34d80       | Neues vom 09. Dezember 2012      | 17                                                 |
|                          | pkop (RhodeCode Admin) | Mon, 10 Dec 2012 05:53:46        | Parent 99:69a2e31516e98<br><a href="#">DETAILS</a> |
| <input type="checkbox"/> | 99:69a2e31516e98       | Link zu Epistel nachgereicht     | 1                                                  |
|                          | pkop (RhodeCode Admin) | Sun, 02 Dec 2012 18:26:03        | Parent 98:3d28166f7b34<br><a href="#">DETAILS</a>  |
| <input type="checkbox"/> | 98:3d28166f7b34        | Neues aus den ersten drei Bänden | 11                                                 |
|                          | pkop (RhodeCode Admin) | Sun, 02 Dec 2012 18:19:11        | Parent 97:b4ebcaabb5aa<br><a href="#">DETAILS</a>  |
| <input type="checkbox"/> | 97:b4ebcaabb5aa        | Neues vom 17. November 2012      | 21                                                 |
|                          | pkop (RhodeCode Admin) | Thu, 17 Nov 2012 19:03:00        | Parent 96:e02a8404d0da<br><a href="#">DETAILS</a>  |

# Kleine Helfer in Python

## Indexerstellung und Seiten mit den gescannten Bildern

```
(rueckert)tinship-2:buecher peterkoppatz$ ls -al
total 40
drwxr-xr-x  7 peterkoppatz  staff   238  9 Mär 22:53 .
drwxr-xr-x  9 peterkoppatz  staff   306  9 Mär 20:29 ..
-rwxr-xr-x  1 peterkoppatz  staff  5576 26 Feb 16:22 Makefile
-rw-r--r--  1 peterkoppatz  staff  1172 26 Feb 16:22 README.txt
-rw-r--r--  1 peterkoppatz  staff  4068 26 Feb 16:22 getIndex.py
-rw-r--r--  1 peterkoppatz  staff  2328 26 Feb 16:22 makeImageFiles.py
drwxr-xr-x 15 peterkoppatz  staff   510 27 Feb 13:26 source
```

# Index

## REST-Format

### Anmerkungen

.. glossary::

Unterschleif  
veraltet: Unterschlagung

``siehe auch Wiktionary <http://de.wiktionary.org/wiki/Unterschleif>`_`

.. index:: Feder  
.. index:: Frau  
.. index:: Geduld  
.. index:: Haushalt  
.. index:: Haushaltsbuch  
.. index:: Liebeslieder  
.. index:: Poesie  
.. index:: Schreiberei  
.. index:: Schreibfeder  
.. index:: Schuld  
.. index:: Spruch  
.. index:: Unterschlagung  
.. index:: Unterschleif  
.. index:: Zahlen

# Index

Spezialindex neben der Volltextsuche

## Index

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O  
| P | Q | R | S | T | U | V | W | Y | Z

### A

---

**Aal**

**Aar**, [1], [2]

**Aare**, [1]

**Abend**, [1], [2], [3], [4],  
[5], [6], [7], [8], [9]

**Abendglanz**

**Abendhimmel**

**Abendmahle**

**Abendroth**, [1], [2]

**Abendröthe**, [1], [2]

**Abends**, [1], [2]

**Abendschimmer**

**Abendsonnenlicht**

**Abendstern**

**Abenteuer**, [1]

**Abermannsdorf**, [1]

**Amorinen**

**Amors**, [1], [2], [3], [4],  
[5]

**Amphions**

**Amt**, [1], [2], [3]

**Amte**, [1]

**Amtes**

**Anachoreten**, [1]

**Anbaus**

**Anbeginne**

**Anblick**, [1], [2], [3], [4]

**Andacht**, [1]

**Anderen**

**Anerkennung**

**Anfang**, [1], [2]

**Andenken**, [1], [2], [3]

# Scan-Vorlage

## Korrekturlesen

Navigation -

Search

### Werke Band 1; Buch 1, Kapitel 5, Seite 253

— 253 —

#### Herbstgefühl.

Wie ein herbstdurchschütterter Strauch  
Ist das zagende Vaterland ;  
Wo in Blättern sich regt ein Hauch,  
Löst er einem das Lebensband.  
Wie das sterbende Blatt sich schmückt,  
Küßt es weinend der Sonnenstrahl ;  
Frühlingstäuschung, die mich beglückt,  
Ach du lächelst zum letzten Mal.  
Vögel fühlen den Winter vor ;  
Wie die wandern im Nebeldunst,  
Senken die sich in Schilf und Rohr,  
Die zum Schlafen in Fels und Klust.  
Glücklich sind die schlafen, und die  
Sind beglückter, die wandern aus.  
Die da wachen und bleiben hie,  
Klagen in Frost und Wintergraus.

PK

- Software, die im Rückert-Projekt benutzt wird
- Wie man helfen kann . . .
- Ich hab schon eine Idee – Kontakt/Infos

# Zusammenfassung

Software, die im Rückert-Projekt benutzt wird

- 1 OCR/Texterkennung inklusive Fraktur-Schrift mit **tesseract**
- 2 Rohtexte im **REST**-Format erstellen und speichern
- 3 Transformation mit **Sphinx**
- 4 Datensicherung mit einem verteilten Versions-Kontroll-System (Git, Bazaar, **Mercurial**)
- 5 Publizieren auf [gutenberg.org](http://gutenberg.org), [openlibrary.org](http://openlibrary.org), **eigener Website**
- 6 Automatisierte Verarbeitung mit einer Sprache der Wahl, z. B. **Python**



## Was man selbst tun kann, um Alternativen anzubieten

- 1 Spenden an Projekte, die Infrastruktur zur Verfügung stellen und unterhalten
- 2 Wer sein Lieblingsbuch anderen zugänglich machen will, ist herzlich eingeladen, es selbst zu digitalisieren.
- 3 Werden Sie Korrekturleser («Distributed Proofreader», <http://www.pgdp.net>)
- 4 Erzählen Sie jedem, dass es solche Projekte gibt.
- 5 Lesen Sie wieder einmal ein Buch, ein Gedicht oder andere Nicht-Fachliteratur 😊.



# Ich hab schon eine Idee

Kontakt/Infos

## Kontakt: Peter Koppatz

Web: <http://rueckert.gesammelte-werke.org>

Web: <http://www.sudile.com>

E-Mail: [pkoppatz@web.de](mailto:pkoppatz@web.de)

Tel: (0)331 200 297 17

### **Postanschrift**

Sudile GbR

Jägerstraße 36

14467 Potsdam



# Nachbetrachtung

Erfahrungsaustausch ist die billigste Investition.

In der Diskussion kamen wichtige Hinweise von Zuhörern, die es zu beachten gilt bzw. zu testen lohnt.

- 1 Bilder mit mindestens 400 DPI scannen
- 2 Schief gescannte Bilder gerade rücken:  
<http://ortyl.org/pgmangle.html>
- 3 Bevorzugte Bildformate für die Weiterverarbeitung durch eine OCR-Software ist PNG oder TIFF, JPG eher nicht verwenden, weil die Kanten weichgezeichnet werden.
- 4 Für bessere Ergebnisse sollte Tesseract «trainiert» werden.