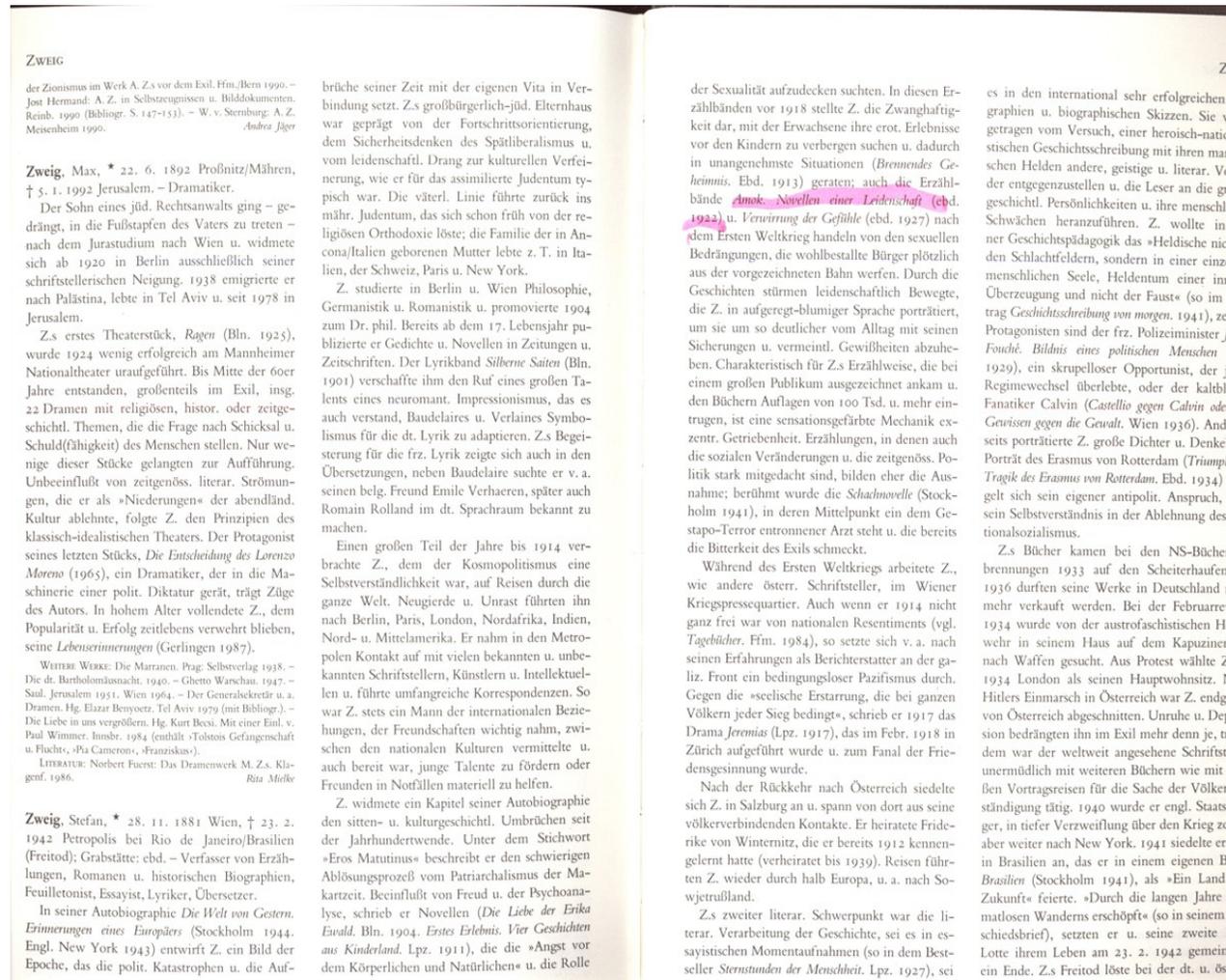


Vom Antiquariat zum E-Text

Das Projekt Gutenberg-DE hat im Laufe von über 20 Jahren 8500 digitale Werke (Stand: Oktober 2016) produziert.

Den Arbeitsablauf und -aufwand soll dieses PDF verdeutlichen.

1. Schritt: Auswahl ...



Aus verschiedensten Quellen werden Ideen für Neuzugänge für Gutenberg-DE gesammelt. Neben einer ganzen Reihe von literaturwissenschaftlichen Werken und Literaturlexika aus dem 19. Jahrhundert, dem Internet und einigen wenigen Leseranregungen werden auch auch zeitgenössische Lexika verwendet. Hier: Walther Killy, LITERATURLEXIKON, Band 12, Vas - Z.

... und Kauf der antiquarischen Buchvorlage

The screenshot shows a web browser window with the ZVAB website. The browser's address bar shows the URL: www.zvab.com/servlet/SearchResults?an=zweig&bi=0&ds=30&exactsearch=off&recentlyadded=all&sortby=20&sts=t&n=Amok. The browser tabs include 'GaGa Startseite -- Gold E-Texte', 'zweig - Amok - ZVAB', and 'Amok'. The website's navigation bar includes links for 'Deutsches Literaturar...', 'DNB Personennamend...', 'index.htm', 'GaGa Startseite', 'Alle gaga-Texte', 'Archiv', 'Gutenberg-DE Scans', 'Gutenberg-DE autoren', and 'Gutenberg-DE Werke'. The main content area displays two search results for the book 'Amok'.

6.

Zweig, Stefan::
Amok.
[nach diesem Titel suchen]

Amok. Novellen einer Leidenschaft. Leipzig, Insel, 1921 317 S. OLn, 1921

Anbieter
Antiquariat Udo Schwörer, (Pforzheim, Deutschland)
Bewertung: ★★★★★
Anzahl: 1

Sprache: Deutsch Gewicht in Gramm: 900. Gut. Artikel-Nr.: 34799.
Weitere Bücher des Anbieters | Frage an den Anbieter

In den Warenkorb

Preis: EUR 14,00
Währung umrechnen

Versand: EUR 3,00
Innerhalb Deutschland
Versandziele, Kosten & Dauer
Anbieter- und Zahlungsinformationen

7.

ZWEIG STEFAN:
Amok.
[nach diesem Titel suchen]

Lpz Insel Verlag, 1927

Anbieter
Neusser Buch & Kunst Antiquariat, (Neuss, NRW, Deutschland)
Bewertung: ★★★★★
Anzahl: 1

46. - 50. Tsd. 8°. 295 S., 2 Bil. OLeinen m. Goldaufschrift a. Deckel u. Rü., leiche Rückenvergoldung, Kopffarbschnitt, Gebrauchsspuren, gut erhalten. 0,500 gr. 50. Artikel-Nr.: 26252AB.
Weitere Bücher des Anbieters | Frage an den Anbieter

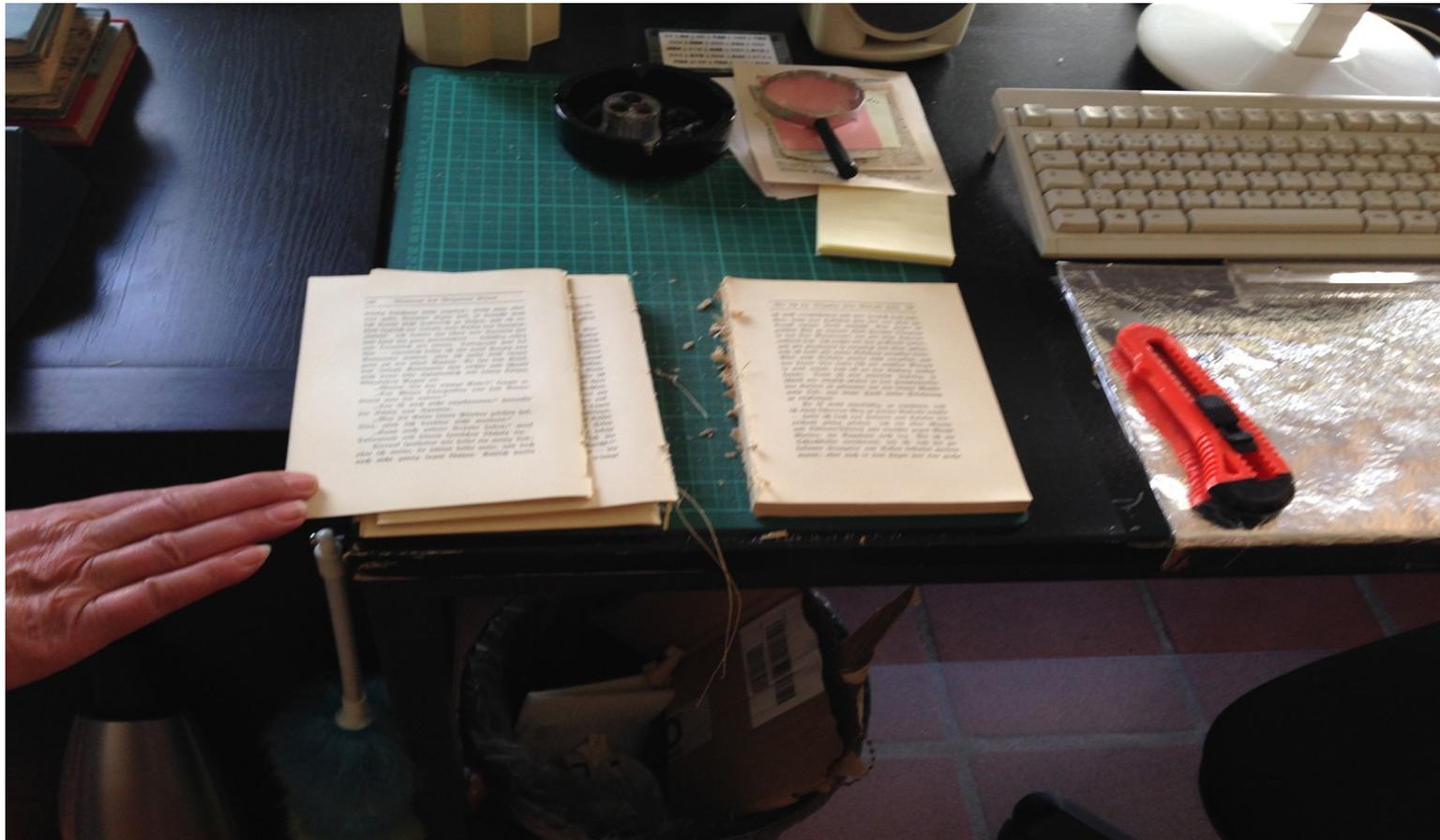
In den Warenkorb

Preis: EUR 16,40
Währung umrechnen

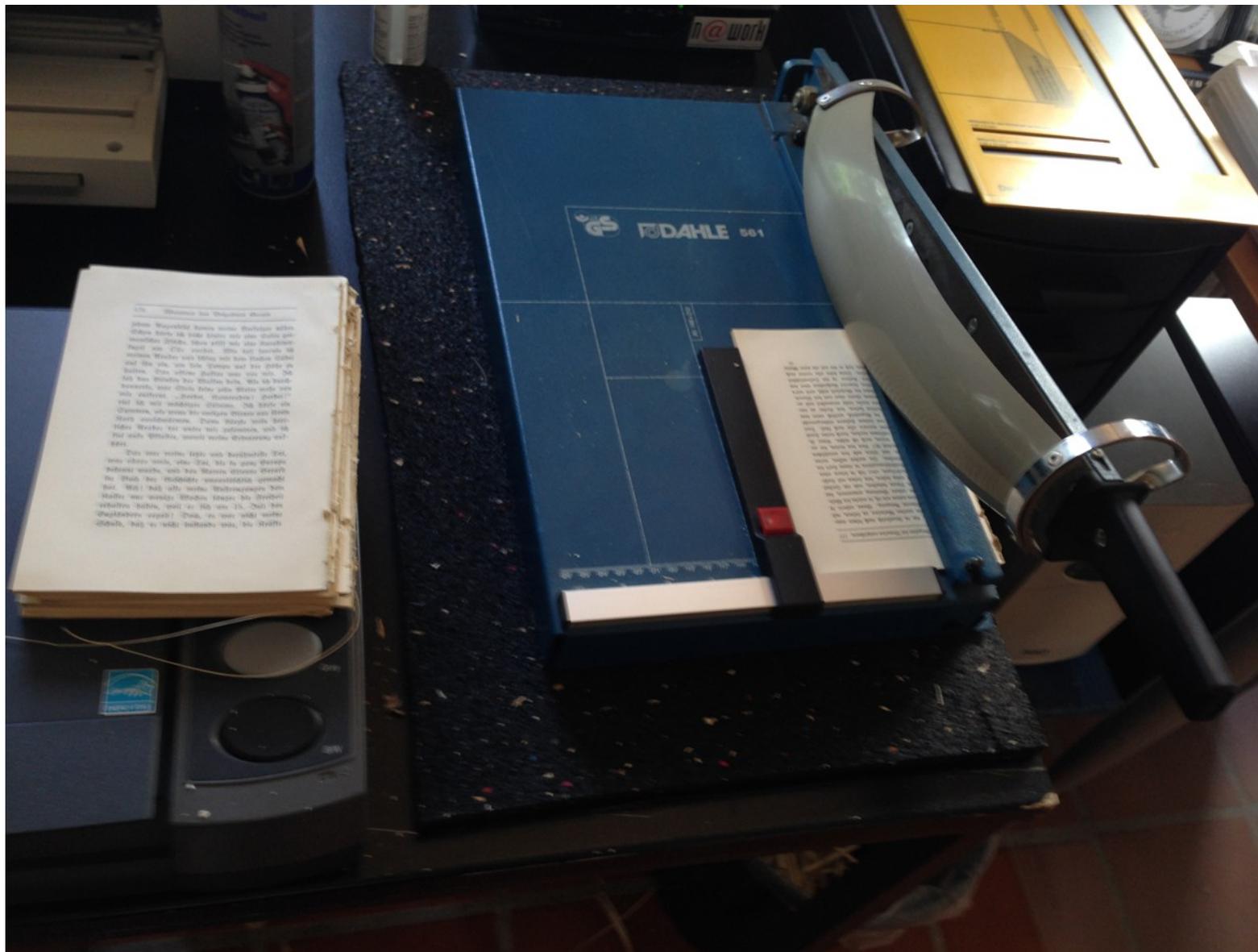
Versand: EUR 3,00
Innerhalb Deutschland
Versandziele, Kosten & Dauer
Anbieter- und Zahlungsinformationen

Gutenberg-DE kauft derzeit ca. 40 Bücher/Monat. Bei einem mittleren Preis von nur € 10,--/Buch ergeben sich Investitionen für die Buchanschaffung von bisher fast hunderttausend €. Zeitaufwand für Schritt 1: 1/2 Stunde/Buch.

2. Schritt: Scannen



Das Buch wird zerteilt ...



... sauber beschnitten und anschließend ...



... in den Scanner gelegt und gescannt. Altersbedingt ist das Papier oft brüchig, dann müssen sorgfältig alle Seiten von Hand eingelegt werden.

Stefan Zweig

Amok

Novellen einer Leidenschaft

1928

Im Insel-Verlag zu Leipzig

Frans Masereel,
dem Künstler, dem brüderlichen Freunde
Salzburg, Frühling 1922

Du auf dich, Unterwelt der Leidenschaften:
Gestalten ihr, geträumt und doch empfunden,
Laßt eure Lippen heiß an meinen haften,
Trinkt Blut von Blut und Atem mir vom Munde!

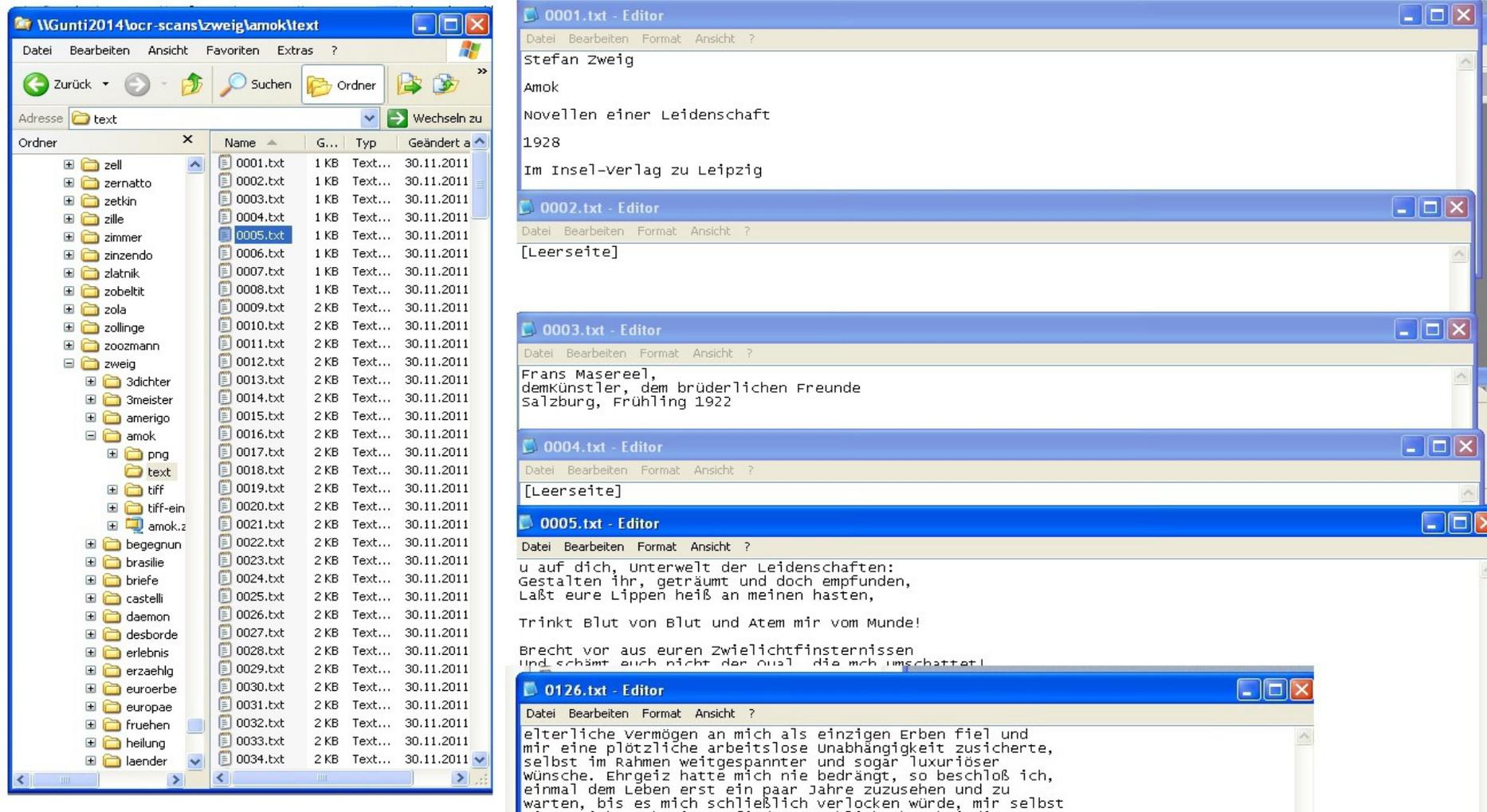
Brecht vor aus euren Zweifelsfinsternissen
Und schämt euch nicht der Dual, die euch umschattet!
Wer Liebe liebt, will nicht ihr Leiden missen,
Was euch verflört, ist's, was mich zu euch gattet.

Nur Leidenschaft, die ihren Abgrund findet,
Läßt deine letzte Wesenheit entbrennen,
Nur der sich ganz verliert, ist sich gegeben.

So flamm dich auf! Erst wenn du dich entzündet,
Wirst du die Welt in deiner Tiefe kennen:
Erst wo Geheimnis wirkt, beginnt das Leben.

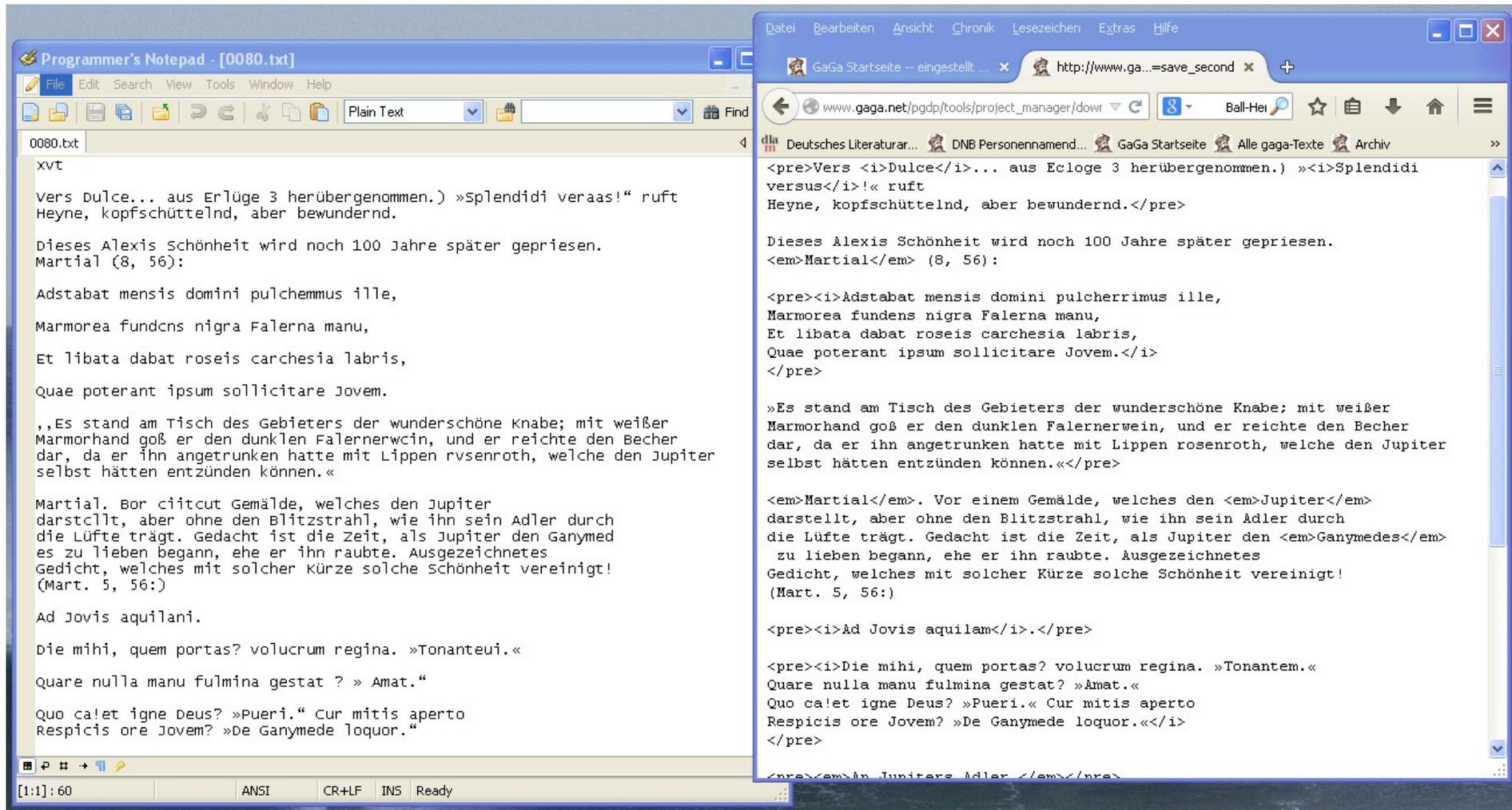
Dies sind die gescannten Bilder der ersten Seiten, die für den Computer, wie für jeden Analphabeten, nicht als ein Muster von schwarzen Punkten auf weißer Fläche sind.

3. Schritt: OCR-Bearbeitung

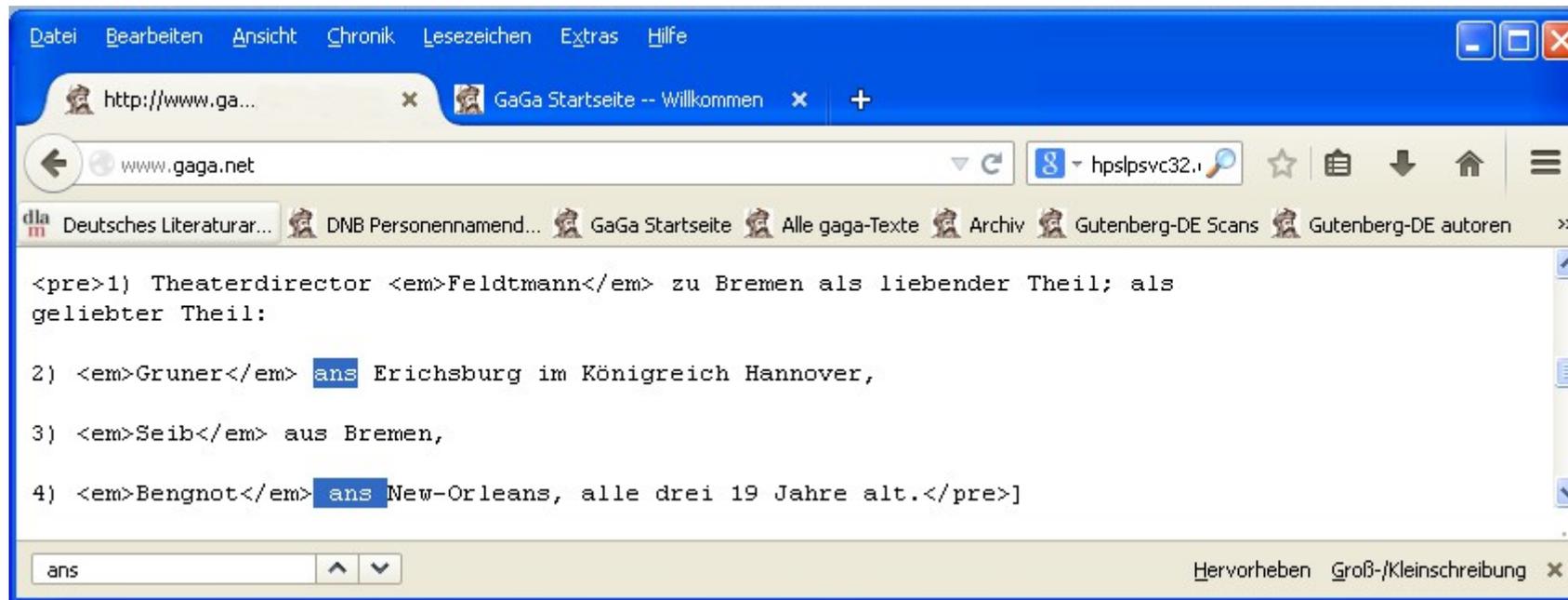


Mittels spezieller Programme (OCR-Programme, OCR = Optical Character Recognition. Eine Software, die ein Muster schwarzer Punkte in Buchstaben umsetzt) wird das Bild der Seite in digitalen Rohtext umgewandelt. Zeitaufwand für Schritt 2 und 3: 3/4 Stunde/Buch

4. Schritt: Korrekturlesen. Die OCR-Software macht Fehler, die nur durch Korrekturlesen gefunden und entfernt werden können. Typische Fehler bei Frakturschrift sind Verwechslungen von f/s, u/n, t/l und andere. Selbst per ausgeklügelter Software ist eine automatische Korrektur dieser Fehler nicht möglich, denn ob es heißt "alte werdende Eltern" oder "alle werdende Eltern", "er genoß die Luft" oder "er genoß die Lust", ob jemand "ausgibt" oder "aufgibt" kann nur der menschliche Leser aus dem Kontext entscheiden. Auch eine Rechtschreibprüfung hilft hier nicht weiter, weil alles richtig geschriebene deutsche Wörter sind. Jede Buchseite wird in zwei aufeinanderfolgenden Runden doppelt unter Abgleichung mit dem gescannten Buch korrekturgelesen. Von den bereits publizierten 8.500 Texten wurden bei www.gaga.net ca. 3.700 Texte korrekturgelesen, 4.800 Texte wurden vom Gutenberg-Team bearbeitet. *Links: Roh-OCR (Text, wie aus der Maschine kommt), rechts: Text nach der 2. Leserunde.*



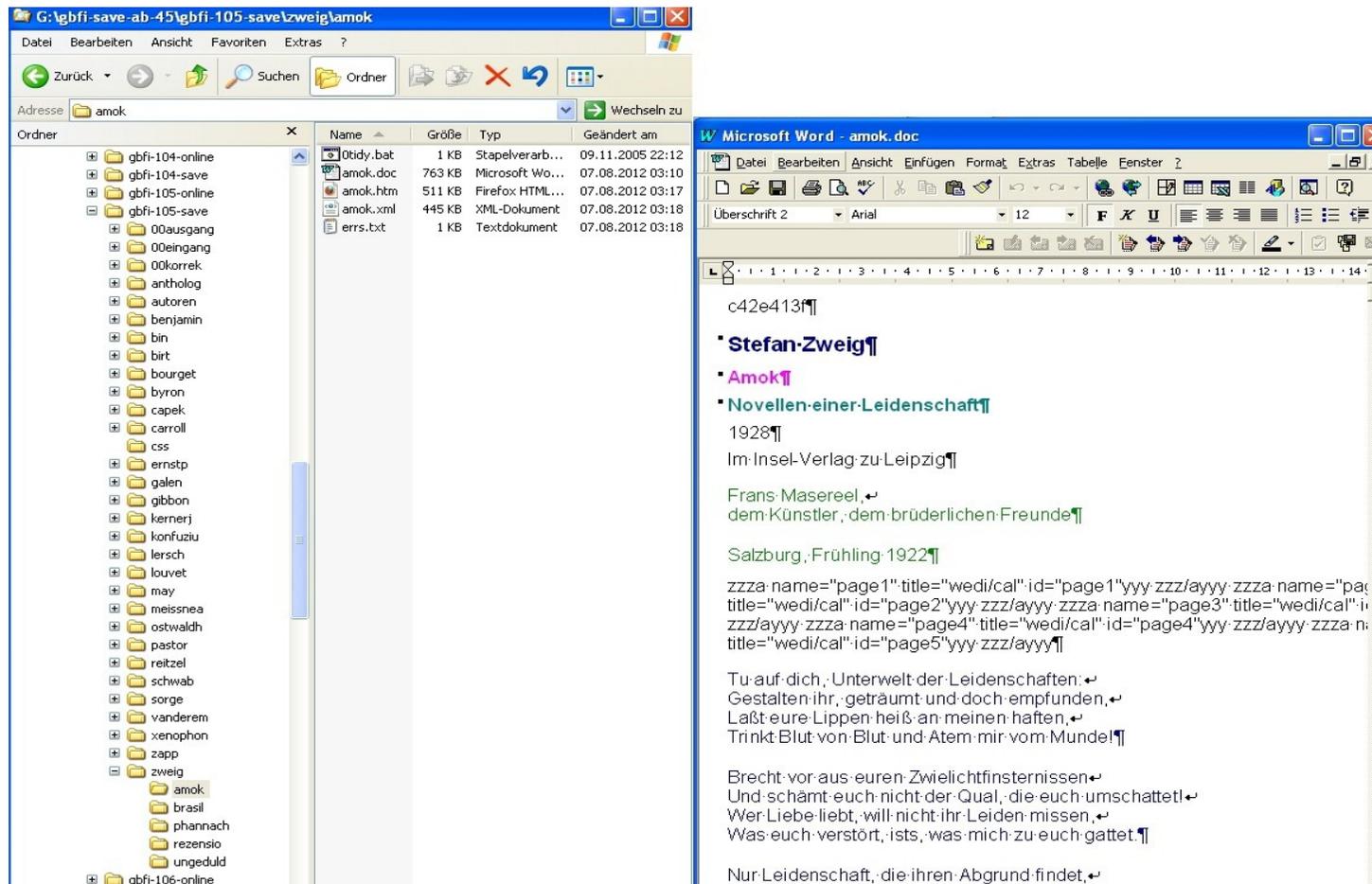
Niemand ist unfehlbar, so werden auch in der zweiten Leserunde immer einige Fehler übersehen. Die meisten dieser Fehler werden in der Nachkorrektur beseitigt. Den tückischen, weil in der Rechtschreibprüfung nicht auffindbaren Rest der Fehler kann man nur eliminieren, wenn das Buch ein weiteres Mal gelesen wird. Hier ein Beispiel: Fehler, die auch in der zweiten Korrekturrunde übersehen wurden.



5. Schritt. Nach dem doppelten Korrekturlesen

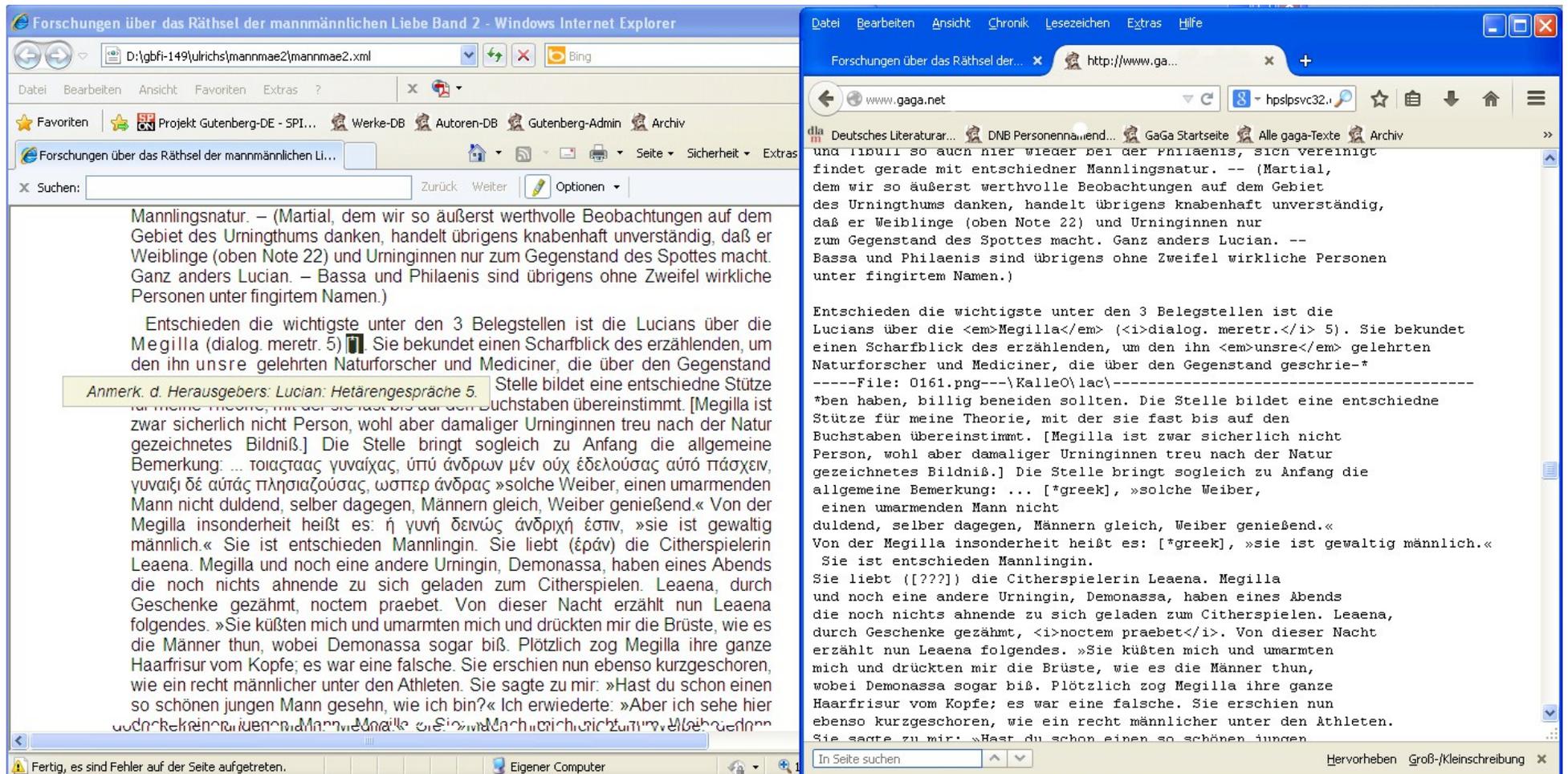
Nach dem Korrekturlesen werden die einzelnen Textdateien zu einem Dokument zusammengefügt, das noch keinerlei Struktur hat. Dies ist der einzige Schritt, der automatisch per Programm erfolgt.

6. Schritt: Nachbearbeitung



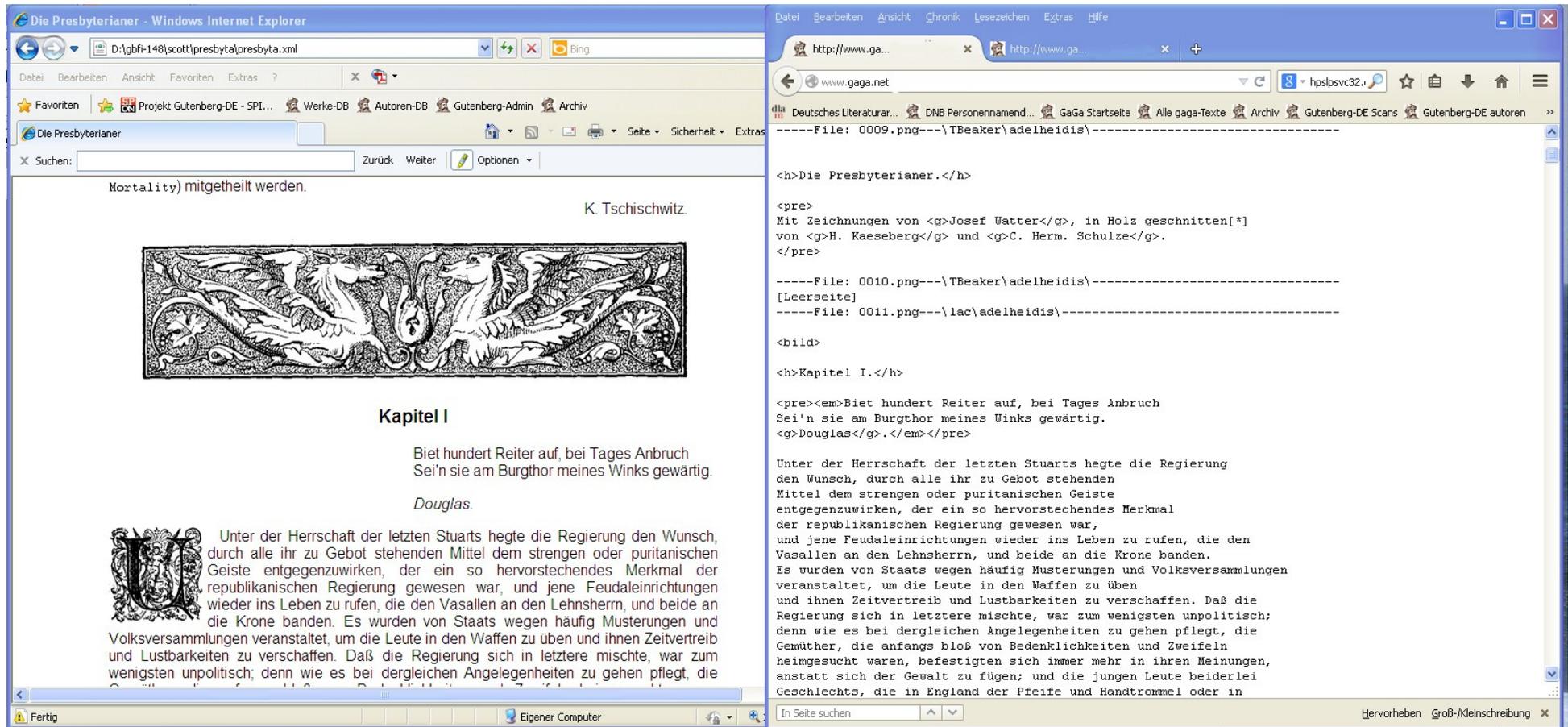
Nachbearbeitung a) Text:

Aus dem doppelt gelesenen Text im vorigen Bild wird ein Worddokument erstellt. Anschließend wird eine Rechtschreibprüfung durchgeführt mittels vieler eigens für verschiedene Sprachepochen erstellter Wörterbücher. Anhand der z.T. gestaffelten Überschriften bekommt der Text eine Struktur. Besonderen Textteilen, wie Gedichten, Briefen, Widmungen u.A. wird eine Formatvorlage zugewiesen. (Die Farben dienen nur zur Kenntlichmachung der jeweiligen Formatvorlagen, sie erscheinen nicht im fertigen Dokument).



Nachbearbeitung b), Sonderzeichen und Fußnoten: (links der fertige Text, rechts der Text nach der zweiten Korrekturrunde)

Aus einem aktuellen Beispiel. Das Buch enthält viel Altgriechisch, das in der Nachbearbeitung eingefügt wird, sowie Anmerkungen des Herausgebers am Ende des Buches, die als Fußnoten an die entsprechende Verweisstelle kopiert und mit dem Zusatz "Anmerkung des Herausgebers" wiedergegeben werden. Dazu kommen noch die Fußnoten des Autors. Unser Rekord liegt bei einem Buch mit 1200 Fußnoten.



Nachbearbeitung c) Bilder: (links der fertige Text, rechts der Text nach 2. Korrekturrunde)

Ungefähr jeder 5. Buch enthält Bilder. Diese müssen nachbearbeitet und Fotos, Farbbilder usw. extra gescannt, in einem Grafikprogramm nachbearbeitet und eingepflegt werden.

Forschungen über das Rätsel der mann männlichen Liebe Band 2 - Windows Internet Explorer

D:\gbfi-149\ulrichs\mannmae2\mannmae2.xml

Projekt Gutenberg-DE - SPI... Werke-DB Autoren-DB Gutenberg-Admin Archiv

Forschungen über das Rätsel der mann männlichen Li...

Suchen: Zurück Weiter Optionen

Genossen zur Herausgabe von Gladiua furens und Memnon vertheilen sich, wie folgt:

Deutschland:	
Bremen	43 <i>msb</i> = 75¼ fl.
Mecklenburg	2 " = 3½ fl.
Halberstadt	4 " = 7 fl.
Berlin (mein Berichterstatter 3 <i>msb</i>)	
ein anderer 20 <i>msb</i>	25 " = 43¾ fl.
ein 3ter 2 <i>msb</i>	
München	4 fl.
Würzburg	5 " = 8¼ fl.
Ein Hamburger in Wien Öst. W. 5 fl.	5 fl.
10 (sic) Wiener 12 fl. öst. W.	12 fl.
Ein Dioning bei Darmstadt	2 fl.
Ausland:	
Ungarn 5 fl. öst. W. i. S.	= 5 5/6 fl.
Schweiz 1 fl.	
ein Hamburger in der Schweiz 5 <i>msb</i>	= 19¼ fl.
ein Deutscher in Basel 10 fl.	
Paris 10 <i>msb</i> , ein Mecklenburger in Paris 6 <i>msb</i>	16 " = 28 fl.
London 5 Pfund Sterling,	
ein zweiter daselbst 5 Pfund,	= 196¼ fl.
mein dortiger Berichterstatter 10 <i>msb</i> und 5 Pfund	
Summe	411 1/12 fl.

Fertig, es sind Fehler auf der Seite aufgetreten. Eigener Computer

http://www.ga... GaGa Startseite -- Willkommen

www.gaga.net

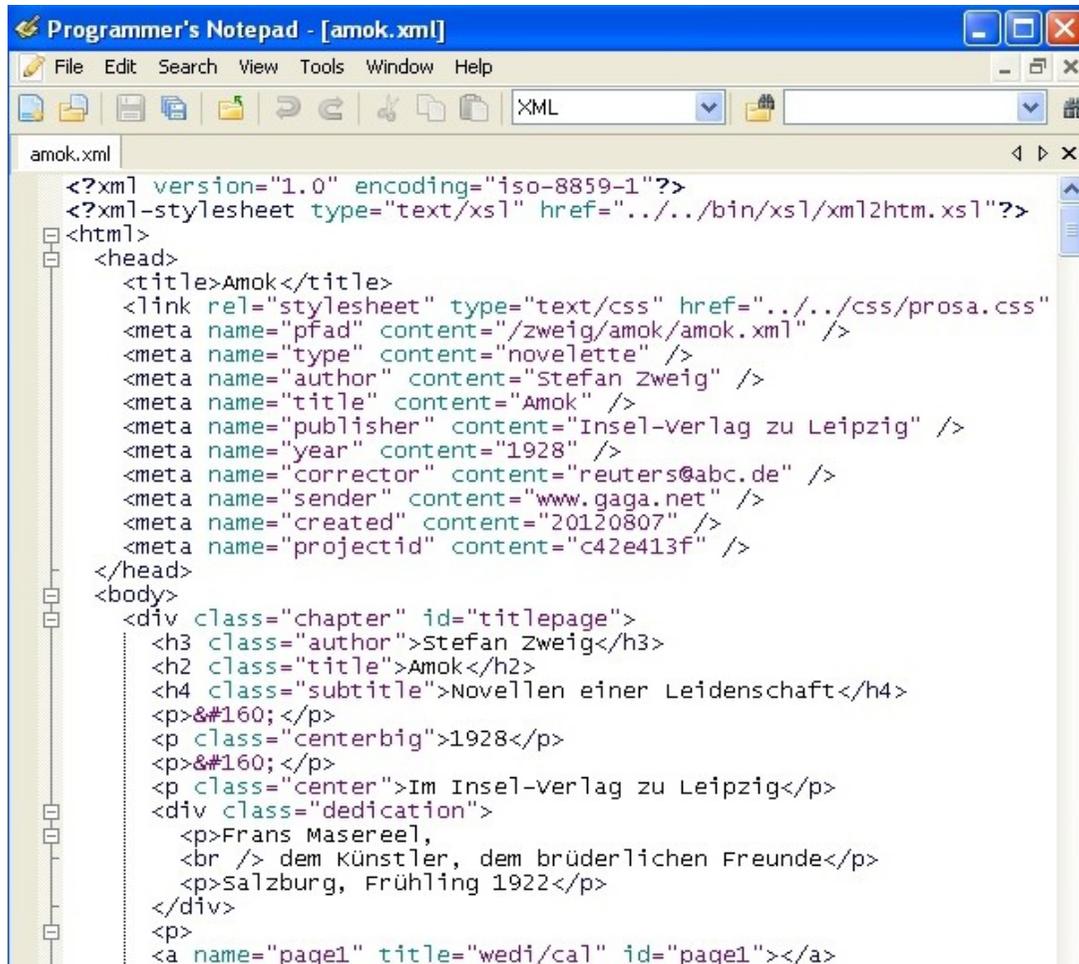
Deutsches Literatur... DNB Personennamend... GaGa Startseite Alle gaga-Texte Archiv Guten...

```
<pre>[*Tabelle*] <em>Deutschland</em>:
Bremen ... 43 [***] = 75¼ fl.
Mecklenburg ... 2 [***] = 3½ fl.
Halberstadt ... 4 [***] = 7 fl.
Berlin (mein Berichterstatter 3 [***], ein
anderer 20 [***], ein 3ter 2 [***]) ... 25 [***] = 43¾ fl.
<hr>
74 [***] = 129¼ fl.</pre>
-----File: 0252.png---\KalleO\MESswein\-----
<pre>Uebertrag 74 [***] -- 129¼ fl.
München ... 4 fl.
Würzburg ... 5 [***] = 8¼ fl.
Ein Hamburger in Wien 5 fl. Öst. W. 5 fl.
Wien, 10 (<i>sic</i>) Wiener 12 fl. öst. W. 12 fl.
Ein Dioning bei Darmstadt .... 2 fl.
<em>Ausland:</em>
Ungarn ... 5 fl. öst. W. i. S. = 5 5/6[*] fl.
Schweiz 1 fl., ein Hamburger in der
Schweiz 5 [***] ein Deutscher in Basel
10 fl ... - ... = 19¼ fl.
Paris 10 [***], ein Mecklenburger in
Paris 6 [***] 16 [***] = 28 fl.
London 5 Pfund Sterling, ein zweiter
daselbst 5 Pfund, mein dortiger Berichterstatter
10 [***] und 5 Pfund . = 196¼ fl.
<hr>
Summe = 411 1/12[*] fl.
</pre>
[*ENDE Tabelle*]
```

Mecklen

Nachbearbeitung d) Tabellen (links der fertige Text, rechts der Text nach 2. Korrekturrunde)

Eine der unbeliebtesten Nachbearbeitungsjobs sind Sachbücher mit vielen Tabellen, da eine elend fummelige Arbeit.

The image shows a screenshot of a text editor window titled "Programmer's Notepad - [amok.xml]". The window has a menu bar with "File", "Edit", "Search", "View", "Tools", "Window", and "Help". Below the menu bar is a toolbar with various icons for file operations and editing. The main text area contains XML code for an HTML document. The code starts with an XML declaration and a stylesheet link. The document structure includes a head section with a title "Amok" and several meta tags for "pfad", "type", "author", "title", "publisher", "year", "corrector", "sender", "created", and "projectid". The body section contains a div with class "chapter" and id "titlepage", which includes an author name, a title, a subtitle, and several paragraphs of text, including a dedication to Frans Masereel.

```
<?xml version="1.0" encoding="iso-8859-1"?>
<?xml-stylesheet type="text/xsl" href="../../bin/xsl/xml2htm.xsl"?>
<html>
<head>
<title>Amok</title>
<link rel="stylesheet" type="text/css" href="../../css/prosa.css"
<meta name="pfad" content="/zweig/amok/amok.xml" />
<meta name="type" content="novelette" />
<meta name="author" content="Stefan Zweig" />
<meta name="title" content="Amok" />
<meta name="publisher" content="Insel-verlag zu Leipzig" />
<meta name="year" content="1928" />
<meta name="corrector" content="reuters@abc.de" />
<meta name="sender" content="www.gaga.net" />
<meta name="created" content="20120807" />
<meta name="projectid" content="c42e413f" />
</head>
<body>
<div class="chapter" id="titlepage">
<h3 class="author">Stefan Zweig</h3>
<h2 class="title">Amok</h2>
<h4 class="subtitle">Novellen einer Leidenschaft</h4>
<p>¶</p>
<p class="centerbig">1928</p>
<p>¶</p>
<p class="center">Im Insel-verlag zu Leipzig</p>
<div class="dedication">
<p>Frans Masereel,
<br /> dem Künstler, dem brüderlichen Freunde</p>
<p>Salzburg, Frühling 1922</p>
</div>
<p>
<a name="page1" title="wedi/cal" id="page1"></a>
```

7. Schritt: Konversion nach XML

Aus dem Worddokument wird eine XML-Datei hergestellt, die als Ausgangsmaterial für alle weiteren Anwendungen archiviert wird. Diese Datei wird zur Internetseite bei Spiegel Online hochgeladen und sie erscheint auf der jeweiligen Edition.

Projekt Gutenberg-DE

Startseite Genres Neue Texte Alle Autoren Alle Werke Lesetips Shop Information Impressum

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

Gutenberg > Stefan Zweig > Amok

Stefan Zweig: Amok - Kapitel 1

Navigation: Kapitel 1

Stefan Zweig
Amok
Novellen einer Leidenschaft

1928

Im Insel-Verlag zu Leipzig

Frans Masereel,
dem Künstler, dem brüderlichen Freunde

Salzburg, Frühling 1922

Tu auf dich, Unterwelt der Leidenschaften:
Gestalten ihr, geträumt und doch empfunden,
Laßt eure Lippen heiß an meinen haften,
Trinkt Blut von Blut und Atem mir vom Munde!

Brecht vor aus euren Zwielftfinsternissen
Und schämt euch nicht der Qual, die euch umschattet!
Wer Liebe liebt, will nicht ihr Leiden missen,
Was euch verstört, ists, was mich zu euch gattet.

pfad /zweig/amok/amok.xml
type novelette
author Stefan Zweig
title Amok
publisher Insel-Verlag zu Leipzig
year 1928
corrector reuters@abc.de
sender www.gaga.net
created 20120807
projectid c42e413f

8. Schritt: Publizieren

Da Gutenberg-DE einen Beitrag zur Förderung der deutschen Sprache und Kultur leisten möchte, soll gute deutsche Literatur von möglichst vielen Menschen gelesen werden. Deshalb werden alle Gutenberg-Texte, mit Quellenangaben versehen, zu Spiegel Online zum kostenlosen lesen hochgeladen. Außerdem geben wir im Abstand von ca. 1 1/2 Jahren den gesamten Textbestand auf Datenträger heraus. Die Einnahmen aus diesen Editionen decken gerade eben die Sachkosten: Miete, Mietnebenkosten, Strom, Geräte, Büromaterial.

9. Schritt: Buchquelle archivieren



Für eine eventuelle spätere Nachkorrekturen, zum Quellennachweis und für die Möglichkeit, Bilder zum Druck hochauflösend nachzuscannen, werden die Vorlagen der fertig digitalisierten Bücher aufbewahrt. Die Bücher werden, da zerschnitten, in Folie eingeschweißt und einsortiert.

(Foto aus dem Büro von Hille & Partner)

Zeitaufwand

Der Zeitaufwand beträgt je digitalisiertem Buch im Mittel:

1. Schritt: Buchauswahl und -kauf	1/2 Stunde
2. und 3. Schritt: Scan und OCR	3/4 Stunde
4. Schritt: Korrekturlesen	40 Stunden
5. Schritt: Textzusammenfassung	automatisch
6. Schritt: Nachkorrektur und Formatierung	2 Stunden
7. Schritt: Konversion nach XML	1/4 Stunde
8. Schritt. publizieren	Nicht Teil dieser Rechnung
<u>9. Schritt: Buchquelle archivieren</u>	<u>Nicht Teil dieser Rechnung</u>
Summe:	43,5 Stunden x 8300 Bücher = ca. 360.000 Stunden

Selbst wenn man großzügig die Hälfte des Zeitaufwandes als von freiwilligen Helfern erbracht abzieht, bleiben 180.000 Arbeitsstunden, die das Gutenberg-Team im Laufe von 22 Jahren erbracht hat.