



Staats- und  
Universitätsbibliothek  
Bremen

# Nachbearbeitung des OCR-Volltextes der Zeitschrift

## *Die Grenzboten*



Staats- und  
Universitätsbibliothek  
Bremen



- Eine prototypische Zusammenarbeit zwischen Bibliothek und Forschungsinfrastrukturen

Gefördert durch die





## // Digitalisierungsprojekt *Die Grenzboten*

### 1. Das Digitalisierungsprojekt „Die Grenzboten“

- Beschreibung der Zeitschrift „Die Grenzboten“
- Projektbeschreibung
- Volltextfassung der Frakturschrift mittels OCR

### 2. Fortsetzungsprojekt „Nachbearbeitung des OCR-Volltextes der Zeitschrift *Die Grenzboten*“

- Motivation und Ziele
- Projektbeschreibung



## // Digitalisierungsprojekt *Die Grenzboten*

### Zeitschrift „Die Grenzboten“ – inhaltlich

- Erschienen: 1841 bis 1922 wöchentlich, zeitweise zweiwöchentlich
- Untertitel ab 1871: "Zeitschrift für Politik
- Gegründet von Ignaz Kuranda (1811-1884) in Brüssel, später in Leipzig und in Berlin verlegt
- Nationalliberale literarisch-politische Zeitschrift

Quelle: Prof. Dr. phil. habil. Holger Böning, „Gutachten zur Bedeutung der Zeitschrift Die Grenzboten“, 12.07.2010

Ein Wort über das deutsche Theater.

(Eine Vorlesung gehalten im Museum zu Frankfurt.)

Studien zur englischen Litteratur der Gegenwart

1890

Humor und Komik in der griechischen Kunst

Der Kampf um die Armee in Österreich-Ungarn

Die Griechen

Eine

französischen Bildhauers.

1903

Der Vertrag über den Pa

Die Satyre.

1841.



# // Digitalisierungsprojekt *Die Grenzboten*

## „Die Grenzboten“ – Digitalisierungsprojekt

- Laufzeit: 18 Monate, 11/2011 – 04/2013
- Mengengerüst:
  - 81 Jahrgänge
  - 270 Bände
  - 33.000 Beiträge,

Eine politische Satyre.

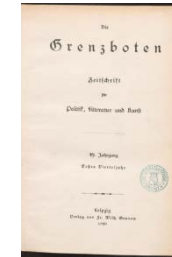
Humor und Komik in der griechischen Kunst

- ca. 185.000 Druckseiten  
~ 3 TB  
(TIFF unkomprimiert)



## Strukturierung und Erschließung

- Strukturierung
  - Setzen der insgesamt 50.000 Strukturknoten mit der Portalverwaltungs-Software (Visual Library Manager, Fa. Semantics)
- Erschließung
  - Manuelle Erfassung der 33.000 Beiträge auf der Grundlage der RAK-WB





# // Digitalisierungsprojekt *Die Grenzboten*

## Präsentation im Internet

- Präsentation im DFG-Viewer

[http://dfg-viewer.de/v2/?set\[mets\]=http%3A//brema.suub.uni-bremen.de%2Foai%2F%3Fverb%3DGetRecord%26metadataPrefix%3Dmets%26identfier%3D282153](http://dfg-viewer.de/v2/?set[mets]=http%3A//brema.suub.uni-bremen.de%2Foai%2F%3Fverb%3DGetRecord%26metadataPrefix%3Dmets%26identfier%3D282153)

sowie im Portal „Digitale Sammlungen der SuUB Bremen“

<http://brema.suub.uni-bremen.de/grenzboten>

- Strukturierte Darstellung
  - Titel, Jahrgänge, Beiträge / Strukturknoten
- Suchfunktionalität mit Highlighting
- PDFs mit Volltext





Staats- und  
Universitätsbibliothek  
Bremen

[Über uns](#) [Service & Beratung](#) [Kataloge & Sammlungen](#) [Fachinformationen](#) [Standorte](#) [Kontakt](#) [Literatur verwalten & publizieren](#)

Home → [Kataloge & Sammlungen](#) → [Digitale Sammlungen](#) → [Sammlung Grenzboten](#)

aaa



Detailsuche

Sammlung Grenzboten ▼

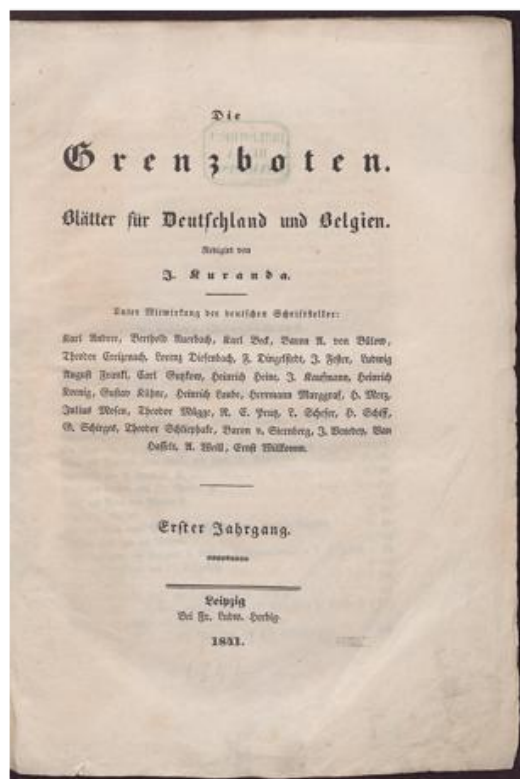
Ihre Suche: z.B. Titel, Thema, Person ...



[Titel](#)

[Inhalt](#)

[Vollbild](#)



In der Zeitschrift suchen

OK

## Titelaufnahme

Titel	Die Grenzboten : Zeitschrift für Politik, Literatur und Kunst
Erschienen	Berlin : Dt. Verl, 1841 - 1922
Online-Ausg.	Berlin : Staatsbibliothek zu Berlin, 2012 - 1999
Anmerkung	Jg. 2.1842 nachträglich als Jg. 1 bez.; 2.1843 auf d. Sem.-Titelbl. fälschlich mit Ersch.-Jahr 1842 angegeben
URN	urn:nbn:de:gbv:46:1-908

## Klassifikation

[Regionalhistorische Werke](#) → [Periodika](#)

[Nicht Regionalspezifisches](#)

## Inhalt

[Inhalt der Zeitschrift](#)

## Metadaten

[METS \(OAI-PMH\)](#)

## Verfügbarkeit

[In meiner Bibliothek](#)

## Nachweis

[Staats- und Universitätsbibliothek Bremen](#)



Staats- und  
Universitätsbibliothek  
Bremen

[Über uns](#) [Service & Beratung](#) [Kataloge & Sammlungen](#) [Fachinformationen](#) [Standorte](#) [Kontakt](#) [Literatur verwalten & publizieren](#)







Home → Kataloge & Sammlungen → Sammlung Grenzboten → Ergebnisliste → Übersicht

aaa  

 Detailsuche  Ihre Suche: z.B. Titel, Thema, Person ... 

[Titel](#) [Inhalt](#) [Übersicht](#) [Seite](#) [Vollbild](#)   1 - 18  

Die Grenzboten > 1850 > 9. Jahrgang. I. Semester. I. B > Vorderdeckel > [1]

 [1]	 [2]	 [3]	 [4]	 [5]	 [6]
 [7]	 [8]	 [9] 1	 [10] 2	 [11] 3	 [12] 4
 [13] 5	 [14] 6	 [15] 7	 [16] 8	 [17] 9	 [18] 10





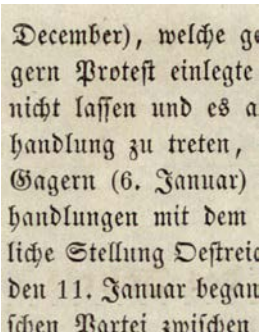
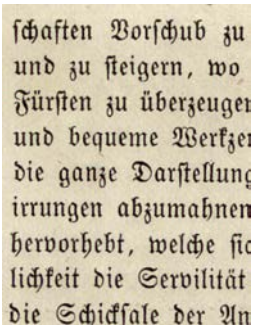
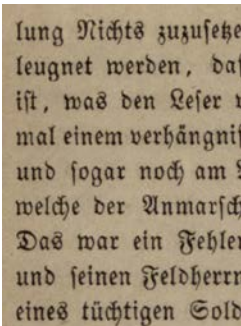
oder die Anfänge solcher Verwirklichung. Nach materialistischer Ansicht werden die Dinge aus ihrem dunkeln Grunde durch den vernunftlosen blinden Drang des Urwesens — was dieses auch sei — hervorgetrieben und nehmen diese oder jene Gestalt an, nicht um einen Zweck zu erfüllen, sondern weil aus der unmittelbar vorhergehenden Lage der Atome keine andre hervorgehen konnte; es giebt in der ganzen Welt kein „damit,“ sondern überall nur ein „weil.“ Beide Ansichten sind, einander bekämpfend, Jahrtausende neben einander hergegangen, und die erste hat die zweite allmählich so weit in sich aufgenommen, daß der Idealist von heute nicht mehr auf einer Schöpfung im Nu besteht, sondern zugiebt, die Welt möge wohl auf dem Wege einer langsamen Entwicklung geworden sein, nur daß es nicht ein blinder Drang des Absoluten oder der Materie, sondern der zwecksetzende, vernünftige Wille des Schöpfers gewesen sei, der diese Entwicklung eingeleitet habe und sie bis zu ihrer Vollendung fortsetzen werde. Im Darwinismus feierte die materialistische Entwicklungstheorie ihren höchsten Triumph, weil durch ihn die Entstehung



# // Digitalisierungsprojekt *Die Grenzboten*

## Volltextgenerierung mittels OCR

- OCR-Software: „FineReader“ (Version 9.0) von der ABBYY Europe GmbH
- Ausgangslage und Vorarbeiten
  - Weitgehend einheitliches, standardisiertes Schriftbild
  - gute Voraussetzungen durch Maschinendruckverfahren, Ende 19. Jhd.

		
1849	1859	1869

- 99,35% Zeichen-Erkennungsquote
- Letzte Bewertung: 98,6% Zeichen-Erkennungsquote

December), welche gegen die Auslegung des ministeriellen Programms durch Gagern Protest einlegte, die Versicherung ertheilte, Oestreich wollte von Deutschland nicht lassen und es ablehnte, **ans** andere Weise mit der **Reichsregierung** in Unterhandlung **zn** treten, als durch den gewöhnlichen Bevollmächtigten. Darauf stellte Gagern (6. Januar) an die Nationalversammlung die Anforderung, ihm **zn** Unterhandlungen mit dem österreichischen Bevollmächtigten über die künftige völkerrechtliche Stellung Oestreichs **zn** Deutschland **Bollmacht** zu **ertheilen**. Die Debatte, welche den 11. Januar begann, zeigte die wachsende Erbitterung, die innerhalb der **dynastischen** Partei zwischen den **Oestreichern** und Preußen ausgebrochen war — denn daß es sich mit dem engern Bundesstaat darum handle, den König von Preußen zum Erbkaiser zu machen, erklärte Gagern damals schon offen, der sich übrigens vergebens bemühte, die Debatte in würdigen und gemessenen Formen **zn** halten. Joseph v. Würth, ehemals **Unterstaatssecretär**, entblödete sich nicht, um Preußen zu schaden, Amtsgeheimnisse auszuplaudern **und** Georg Freiherr von Linke ließ sich dadurch zu einem leidenschaftlichen Ausfall gegen die Oestreicher hinreißen. Schmerling konnte über die eigentlichen Absichten der österreichischen Regierung keine

Auskunft geben **und** so mußte die Nationalversammlung dem Ministerium Gagern (13. Januar) die verlangte Ermächtigung zu Unterhandlungen mit dem **Kaiserstaat** ertheilen. Es geschah mit 261 gegen 224 Stimmen.

Die Versammlung ging darauf an die erste Lesung des Verfassungs-Entwurfs, Die Majorität des Ausschusses stellte den Antrag: die Würde des Reichs-Oberhauptes wird einem der regierenden deutschen Fürsten übertragen. Er wurde von zwei Seiten bekämpft: von der Linken, die den Antrag machte, einen Präsidenten an die Spitze zu stellen, zu welcher Würde jeder Deutsche wählbar sei, und von der Rechten, aus welcher ein Theil ein **Directorium** wollte (Rotenhan), ein anderer (Welcker) einen sechsjährigen Turnus zwischen Oestreich **und** Preußen. Der Antrag der Linken fiel mit 122 : 339 Stimmen; das Amendement Rotenhan mit 97 : 361, das Amendement Welcker mit 80 : 377 **und** so ging der Majoritäts-Antrag mit 258 : 211 Stimmen durch. (19. Januar). Dagegen wurde (23. Januar) die Erbllichkeit des **Reichs-Oberhauptes** verworfen mit 211 : 263, ebenso die Wahl auf Lebenszeit mit 39 : 413, auf sechs Jahre mit 196 : 264, auf drei Jahre 120 : 305, und so blieb die Dauer der neuzuschaffenden Centralgewalt, welcher (25. Januar) der Kaisertitel mit 217 : 205 übertragen wurde, in der Schwebe, wenn man nicht mit Herrn Vogt annehmen wollte, sie sollen auf Kündigung engagirt werde«.

Die kleinen Staaten hatten sich mittlerweile mehrfach für die Uebertragung der **Reichsgewalt** **an** Preußen ausgesprochen; so die Stände von Kassel (5. Januar), Schwerin (9. Januar), Braunschweig (19. Januar) u. s. w. Sehr bestimmt sprachen sich aber die radikalen Kammern von Sachsen, Baiern und Württemberg dagegen aus. Oestreich schlug den geläufigen diplomatischen Weg ein. Es erließ

⇒ auf / **Reichsregierung**

⇒ zu

⇒ zu

⇒ zu / Vollmacht /  
ertheilen

⇒ dynastischen /  
**Oestreichern**

⇒ zu

⇒ **Unterstaatssecretär**

⇒ und

⇒ und

⇒ **Kaiserstaat**

⇒ Directorium

⇒ und

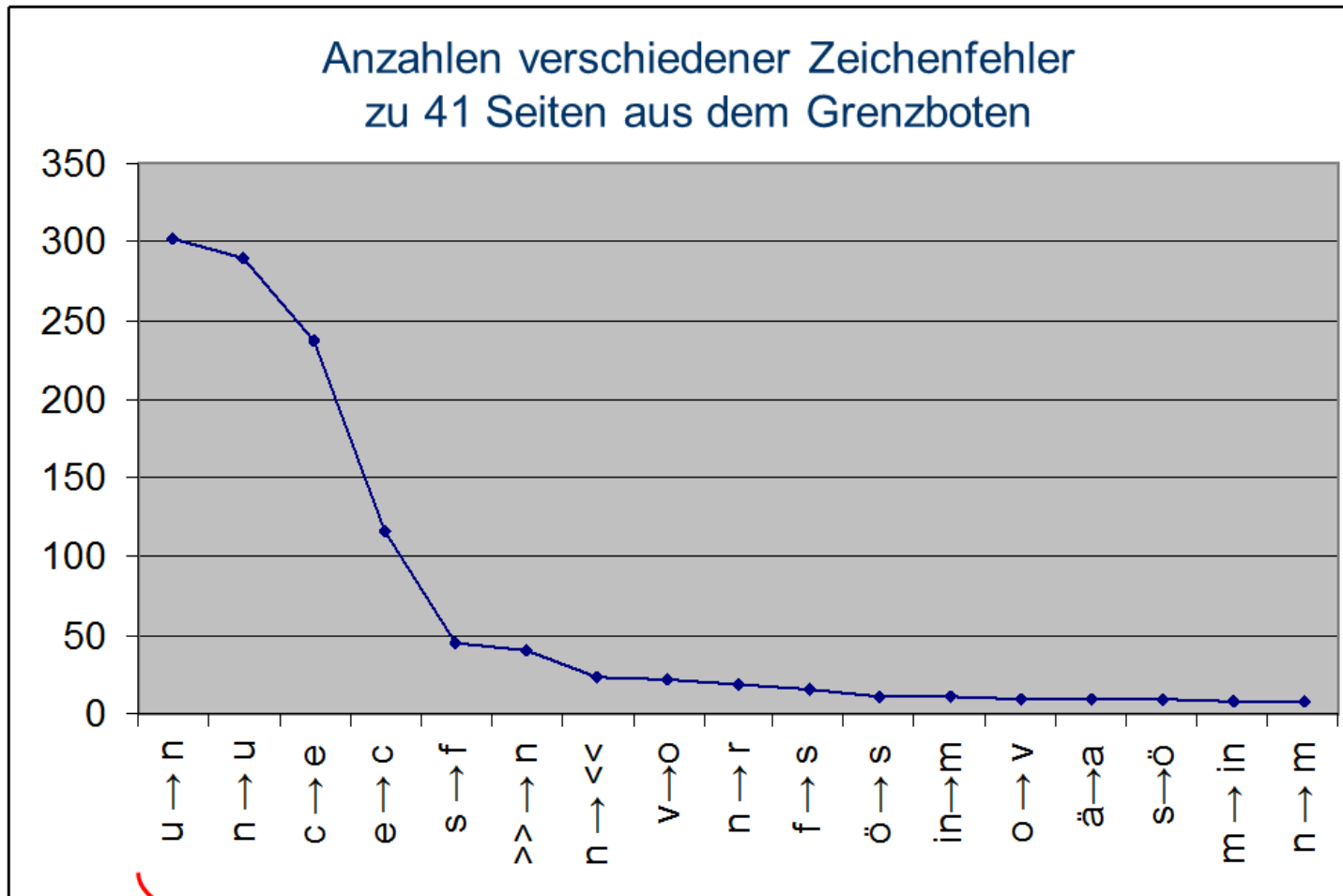
⇒ und

⇒ **Reichs-Oberhaupt**

⇒ **Reichsgewalt** / an



## // Digitalisierungsprojekt *Die Grenzboten*



Summe aller Fehleranzahlen: 1401



## // *Die Grenzboten* – Fortsetzungsprojekt

### 1. Das Digitalisierungsprojekt „Die Grenzboten“

- Beschreibung der Zeitschrift „Die Grenzboten“
- Projektbeschreibung
- Volltexterfassung der Frakturschrift mittels OCR

### 2. Fortsetzungsprojekt „Nachbearbeitung des OCR-Volltextes der Zeitschrift *Die Grenzboten*“

- Motivation und Ziele
- Projektbeschreibung





# // Die Grenzboten – Fortsetzungsprojekt

## Motivation und Ziele

- Verbesserung der Volltextqualität / OCR-Volltext-Nachbearbeitung
  - Zeichenerkennungsquote von aktuell 98,6% verbessern auf 99,5% (Reduktion der Zeichenfehlerquote um 64%)
  - Einsatz [halb]automatisierter Methoden zur Nachbearbeitung von OCR-Volltexten
  - Bewertung aktueller Tools  
→ IMPACT – digitisation.eu – Centre of Competence
- Integration des Grenzboten-Korpus in CLARIN-D / DTAQ
- **Erfahrungsbericht:** Erfassung und Dokumentation der Aufwand-/Nutzen-Verhältnisse beim Einsatz der verschiedenen Ansätze in der bibliothekarischen Praxis



# // Die Grenzboten – Fortsetzungsprojekt

## Konsortium





# // *Die Grenzboten* – Fortsetzungsprojekt

## Aufgaben / Ansätze

- Deliverable 1: **OCR-Volltext-Nachbearbeitung**
  - Vergleich unterschiedlicher Softwaresysteme für Fraktur-OCR
  - Lückenidentifikation und -ergänzung auf der Ebene einzelner Seiten, Absätze oder Zeilen



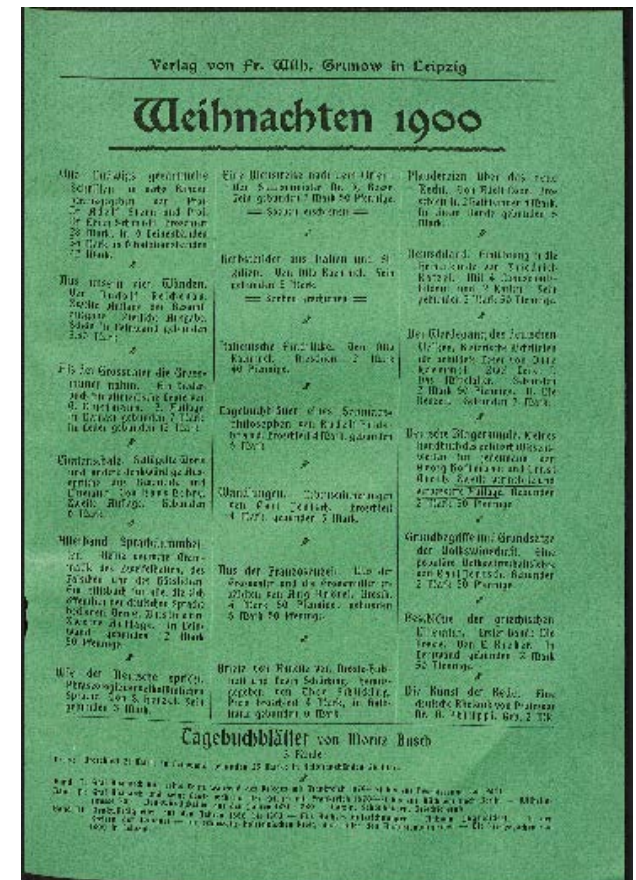
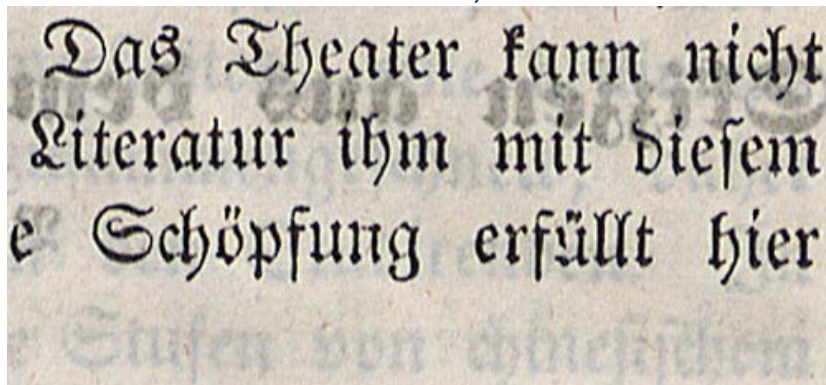
## // *Die Grenzboten* – Fortsetzungsprojekt

### **Lückenidentifikation und -ergänzung auf der Ebene einzelner Seiten, Absätze oder Zeilen**

- Lückenidentifikation
  - Ansatz: Abgleich mit einer Liste historischer Wortformen aus dem DTA-Kernkorpus
    - ca. 3,2 Mio. Wortformen

## Lückenidentifikation und -ergänzung auf der Ebene einzelner Seiten, Absätze oder Zeilen

- Lückenidentifikation
  - Ansatz: Abgleich mit einer Liste historischer Wortformen aus dem DTA-Kernkorpus
    - ca. 3,2 Mio. Wortformen
- Gefunden werden Abschnitte mit
  - Antiqua-Schrift
  - kleiner Schrift
  - Fremdsprachen und Dialekte (lateinisch, französisch, plattdeutsch, ...)
  - kritischen Images: Buchfalz, Kontrast, Widerdruckschatten, ...







# // Die Grenzboten – Fortsetzungsprojekt

## Aufgaben / Ansätze

- Deliverable 1: **OCR-Volltext-Nachbearbeitung**
  - Vergleich unterschiedlicher Softwaresysteme für Fraktur-OCR
  - Lückenidentifikation und -ergänzung auf der Ebene einzelner Seiten, Absätze oder Zeilen
  - Nachkorrektur von OCR-Volltext über Crowdsourcing
    - 1. Wikisource-Treffen in Leipzig am 18./19. Oktober 2013
    - Vereinbarung: Nachkorrektur des 1. Jahrgangs 1841
    - Der Text aus Wikisource bzw. DTAQ wird zur Optimierung von Parametern und Heuristiken für die Volltextkorrektur verwendet



# // Die Grenzboten – Fortsetzungsprojekt

## Aufgaben / Ansätze

- Deliverable 1: **OCR-Volltext-Nachbearbeitung**
  - Vergleich unterschiedlicher Softwaresysteme für Fraktur-OCR
  - Lückenidentifikation und -ergänzung auf der Ebene einzelner Seiten, Absätze oder Zeilen
  - Nachkorrektur von OCR-Volltext über Crowdsourcing

Die Grenzboten (1841/1842) – Wikisource

1. Jg., Band 1, 1841 [Bearbeiten]

**Inhaltsverzeichnis S. 1-8**

- I. Kurnde: Deutschland und Belgien S. 1-10
- D. L. I.: Die belgische Revolution S. 11-15
- A. Baron: Literarische Übersichten vom Standpunkte der Gesellschaft S. 16-25
- Anonymus: Lebzelter Skizzen 1841 S. 26-29
- Tagelbuch S. 30-36
- P. Paris S. 36-37
- L. Bruns: S. 37-38
- Anonymus: Götz 2. Export als Flandern: Selbstbilder auf der Elisenbahn: Gemalt als Mathias: Gernius: Gutknecht: Ein gerechter Urteil: Laube als Dramatiker: Gute Nachrichten aus Stuttgart S. 33-36
- J. F. Schlegel: Ein Wort über das deutsche Theater S. 40-50
- Sch. Elend und Flandern S. 51-55
- A. Baron: Literarische Übersichten vom Standpunkte der Gesellschaft S. 56-59
- B. R.: Briefe aus London I S. 60-64
- P. Paris: Briefe aus Paris S. 65-68
- Tagelbuch S. 67-68
- E. R.: Deutsche Eisenbahnen und belgische S. 69-77
- Alexander von Württemberg: Mohrenkrone S. 78-81
- Anonymus: Alphonse Kann und die Heppen S. 82-85
- Ludwig August Frankl: Eine Begegnung mit Paganini in Genoa S. 86-90
- C. Brille aus Frankfurt I S. 91-99
- Witzler S. 91-99
- I. Kurnde: Bäume in der helvetischen Zeit S. 101-112
- J. G.: Zur Württemberg der neuesten nationalen Bewegungen in Deutschland S. 113-117
- A. Baron: Literarische Übersichten vom Standpunkte der Gesellschaft S. 118-123
- Jobard: Ueber den Bau einer neuen Häuser in Belgien S. 124-127
- A. Weil: Reisebriefe I S. 128-130
- Tagelbuch S. 131-132
- P. Schlegel und Schlegel: S. 133-140
- Anonymus: Lebzelter S. 141-151
- A. Baron: Literarische Übersichten vom Standpunkte der Gesellschaft S. 142-143
- Anonymus: Carl de Brucke über die industriellen Klagen Belgiens S. 153-158
- Literarische S. 159-161
- Tagelbuch S. 162-164
- I. R. (da = Ignaz Kurnde): Auch ein Wort über die vorangegangene Versammlung S. 165-168
- Th. Schlegel: Die Abbildung Kaiser Heinrich des Dritten S. 169-180
- Anonymus: Eine politische Satire S. 181-184
- Tagelbuch S. 185-196
- Th. Schlegel: Deutsche Taschenbücher I S. 197-204

**Taschenbuch**

Autor:	Verschiedene
Titel:	Die Grenzboten
Unterstit:	Zeitung für Deutschland und Belgien, später: Zeitschrift für Politik, Literatur und Kunst
Herausgeber:	Ignaz Kurnde, Gustav Freytag, Louis Schlegel, Georg Clemen
Erscheinungsdatum:	1841-1842
Verlag:	Neufing
Ercheinungsart:	Wochenschrift
Quelle:	Band 1: DDB, Bremen & Commons
Kurzbeschreibung:	1. Jahrgang (in drei Bänden)
Bild:	
Bearbeitungsstand:	unvollständig



Hauptseite  
Systematischer  
Einstieg  
Themenübersicht  
Autorenindex  
Zufällige Seite

Mitmachen  
Letzte Änderungen  
Neuer Artikel  
Korrekturen des  
Monats  
Gemeinschaftsportal  
Skriptorium  
Auskunft  
Hilfe  
Spenden  
Drucken/exportieren  
Buch erstellen  
Als PDF  
herunterladen  
Druckversion

Werkzeuge  
Links auf diese  
Seite  
Änderungen an  
verlinkten Seiten  
Datei hochladen  
Spezialseiten  
dauerhafter Link  
Seiteninformation  
en

Sprachen

Benutzerkonto erstellen Anmelden

Quellentext Diskussion Lesen Bearbeiten Versionsgeschichte Suchen

# Die Grenzboten (1841/1842)

## 1. Jg., Band 1, 1841 [Bearbeiten]

### Seitenindex

- Titel. S. i
- Inhaltsverzeichnis. S. I-II
- I. Kuranda: *Deutschland und Belgien*. S. 1–10
- D. e. t.: *Die Seebäder in Ostende*. S. 11–18
- A. Baron: *Literarische Uebersichten vom Standpuncte der Gesellschaft 1*. S. 19–25
- Anonymus: *Leipziger Skizzen 1841*. S. 26–29
- Tagebuch*. S. 30–36
  - P.: *Paris*. S. 30–31
  - L.: *Brüssel*. S. 31–33
    - Anonymus: *Göthe's Egmont als Flamänder. Selbstmörder auf der Eisenbahn. Geimüller als Maltheser. Gervinus. Gutzkow. Lewald. Ein gerechtes Urtheil. Laube als Dramatiker. Gute Nachrichten aus Stuttgart*. S. 33–36
- J. F.: *September-Scenen*. S. 37–45
- Theodor Greizenach: *Ein Wort über das deutsche Theater*. S. 46–50
- Sch.: *Elsaß und Flandern*. S. 51–55
- A. Weil (=Alexander Weill): *Reisebriefe 1–2*. S. 56–59
- B. B.: *Briefe aus London 1*. S. 60–64
- P.: *Briefe aus Paris*. S. 65–66
- Tagebuch*. S. 67–68
- E. R.: *Deutsche Eisenbahnen und belgische*. S. 69–77
- Alexander von Württemberg: *Mohrenrache*. S. 78–81
- Anonymus: *Alphons Karr und die Wespen*. S. 82–93
- Ludwig August Frankl: *Eine Begegnung mit Paganini in Genua*. S. 94–96
- C.: *Briefe aus Frankfurt 1*. S. 97–99
- Notizen*. S. 91–91
- I. Kuranda: *Börne in der letzten Zeit*. S. 101–112
- J. Cr.: *Zur Würdigung der neuesten nationalen Bewegungen in Deutschland*. S. 113–117
- A. Baron [?]: *Literarische Uebersichten vom Standpuncte der Gesellschaft 2*. S. 118–123
- Jobard: *Ueber den Bau eiserner Häuser in Belgien*. S. 124–127
- A. Weill: *Reisebriefe 3*. S. 128–130
- Tagebuch*. S. 131–132
- P.: *Schauspiel und Schauspieler*. S. 133–140
- Anonymus: *Lacordaire*. S. 141–151
- A. Lebermuth: *Lelewel auf dem Marktplatze zu Brüssel*. S. 143–143
- Anonymus: *Carl de Brouckère über die industriellen Wünsche Belgiens*. S. 153–158
- Literatur*. S. 159–161
- Tagebuch*. S. 162–164
- I. K....da (= Ignaz Kuranda): *Auch ein Wort über die orangistische Verschwörung*. S. 165–168
- Th. Schliephake: *Die Abdankung Kaiser Heinrichs des Vierten*. S. 169–185
- Anonymus: *Eine politische Satyre*. S. 186–194
- Tagebuch*. S. 195–196
- Th. Schliephake: *Deutsche Taschenbücher 1*. S. 197–204

Textdaten	
Autor:	Verschiedene
Titel:	<b>Die Grenzboten</b>
Untertitel:	<i>Blätter für Deutschland und Belgien, später: Zeitschrift für Politik, Literatur und Kunst</i>
Herausgeber:	Ignaz Kuranda, Gustav Freytag, Julian Schmidt, Georg Cleinow
Erscheinungsdatum:	1841–1842
Verlag:	Herbig
Erscheinungsort:	Leipzig
Quelle:	Band 1: SUUB Bremen <span></span> = <span></span> Commons
Kurzbeschreibung:	1. Jahrgang (in drei Bänden)

Artikel in der Wikipedia

Bild

Bearbeitungsstand

unvollständig



# // *Die Grenzboten* – Fortsetzungsprojekt

## Aufgaben / Ansätze

- Deliverable 1: **OCR-Volltext-Nachbearbeitung**
  - Vergleich unterschiedlicher Softwaresysteme für Fraktur-OCR
  - Lückenidentifikation und -ergänzung auf der Ebene einzelner Seiten, Absätze oder Zeilen
  - Nachkorrektur von OCR-Volltext über Crowdsourcing
  - OCR-Volltext-Nachbearbeitung



## // Die Grenzboten – Fortsetzungsprojekt

### OCR-Volltext-Nachbearbeitung

- Korrekturpotential durch die **Liste historischer Wortformen** zusammen mit einer an **Levenshtein** angelehnten **Editierdistanz**

WORT IM GRENZBOTEN	NÄHERUNG AUS DER LISTE	DISTANZ
Scmper	Semper	1
Fürstcnberg	Fürstenberg	1
Municipalvcrfassung	Municipalverfassung	1
Märleiu	Märlein	1
Baucrnfeld	Bauernfeld	1
allern	allem	1
Aktiengcsellschast	Aktiengesellschaft	2
iudividuelleu	individuellen	2
Besttzverhältnifse	Besitzverhältnisse	101
Thevklymenvs	Theoklymenus	200





# // Die Grenzboten – Fortsetzungsprojekt

ocr-visualizer   Home

Fehlerart	Fehler	Anzahl
substitution	[1 1]:O->G[1 1]:n->u[1 1]:n->u[1 1]:n->u[1 1]:n->u[1 4]:n->»[1 1]:u->n[1 1]:n->u[1 1]:e->c[1 1]:n->u[1 1]:u->n[1 1]:i->t[1 1]:u->n[1 1]:u->n[1 1]:u->n[1 1]:e->c[1 1]:n->u[1 1]:n->u[1 1]:n->u[1 1]:s->S[1 1]:n->u[1 1]:u->n[1 1]:s->S[1 1]:f->s[1 1]:o->v[1 1]:e->c[1 1]:n->u[1 1]:s->S	28
many-to-one	[7 1]: ->n[7 1]: l->j	2
deletion	[1 8]:l-> [1 8]:l-> [5 8]: -> [5 8]: -> [5 8]: -> [5 8]: -> [4 8]: -> [4 8]: -> [4 8]: ->	9

434 ■effentliche Charaktere. ■ Johann Jacoby. Es ist noch keine zehn Jahr, als man in Deutschland Ostpreußen wie ein Stück Vorsibirien ansah, wo die neun Mouate lang ■ nicht im Stande sein sollte, den Schnee zu schmelzen, wo die Wölfe schaarenweis in den Straßen der Hauptstadt den Mond anheulten und wo in den wüsten Kieferwäldungen die melancholischen Ele<sup>■</sup>thiere

Man ■hörte ■mir von Zeit zu ■Zeit

Man ■hörte ■nur von Zeit zu ■Zeit

Anacharsis aus diesem modernen Scythien nach Deutschland, um sich zu wärmen und Europas übertünchte Höflichkeit zu studieren: Herder, der Gelehrterath in Bückeburg wurde und nachher sich den Sternreihen der

VLID 277873



# // Die Grenzboten – Fortsetzungsprojekt

## Aufgaben / Ansätze

- Deliverable 1: **OCR-Volltext-Nachbearbeitung**
  - Vergleich unterschiedlicher Softwaresysteme für Fraktur-OCR
  - Lückenidentifikation und -ergänzung auf der Ebene einzelner Seiten, Absätze oder Zeilen
  - Nachkorrektur von OCR-Volltext über Crowdsourcing
  - OCR-Volltext-Nachbearbeitung
  - Bewertung von Tools aus dem Projekt IMPACT – Improving Access to Text





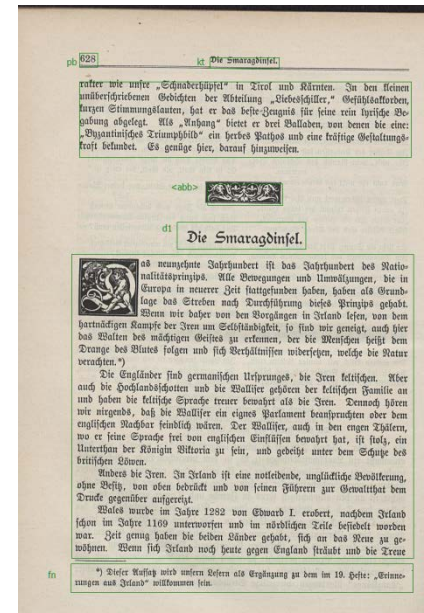
# // *Die Grenzboten* – Fortsetzungsprojekt

## Aufgaben / Ansätze

- Deliverable 1: **OCR-Volltext-Nachbearbeitung**
  - Vergleich unterschiedlicher Softwaresysteme für Fraktur-OCR
  - Lückenidentifikation und -ergänzung auf der Ebene einzelner Seiten, Absätze oder Zeilen
  - Nachkorrektur von OCR-Volltext über Crowdsourcing
  - OCR-Volltext-Nachbearbeitung
  - Bewertung von Tools aus dem Projekt IMPACT – Improving Access to Text
- Deliverable 2: Integration des **Grenzboten-Korpus in CLARIN-D / DTAQ**
  - „Zoning“: Erfassung der textformalen Struktur der gedruckten Vorlage

## „Zoning“ – Erfassung der textformalen Struktur

- ... bis zur Absatzebene
- Die semantischen Funktion wird annotiert → es ist erkennbar, ob es sich um Titelzeile, Fließtext oder Fußnotentext handelt
- Grundlage für die Annotation: DTA-Basisformat / TEI-P5-Richtlinien
- Es wird das vom DTA entwickelte „Zoning“-Tool (DTA-ZOT) weiterentwickelt und eingesetzt



rakter wie unsere „Schnaderhüpfel“ in Tirol und Kärnten. In den kleinen unüberschriebenen Gedichten der Abteilung „Liebeschiller,“ Gefühlsaktorden, kurzen Stimmungslauten, hat er das beste Zeugnis für seine rein lyrische Begabung abgelegt. Als „Anhang“ bietet er drei Balladen, von denen die eine: „Byzantinisches Triumphbild“ ein herbes Pathos und eine kräftige Gestaltungskraft bekundet. Es genüge hier, darauf hinzuweisen.

&lt;abb&gt;



d1

## Die Smaragdinsel.



Das neunzehnte Jahrhundert ist das Jahrhundert des Nationalitätsprinzips. Alle Bewegungen und Umwälzungen, die in Europa in neuerer Zeit stattgefunden haben, haben als Grundlage das Streben nach Durchführung dieses Prinzips gehabt. Wenn wir daher von den Vorgängen in Irland lesen, von dem hartnäckigen Kampfe der Iren um Selbständigkeit, so sind wir geneigt, auch hier das Walten des mächtigen Geistes zu erkennen, der die Menschen heißt dem Drange des Blutes folgen und sich Verhältnissen widersetzen, welche die Natur verachten. \*)

Die Engländer sind germanischen Ursprunges, die Iren keltischen. Aber auch die Hochlandsschotten und die Walliser gehören der keltischen Familie an und haben die keltische Sprache treuer bewahrt als die Iren. Dennoch hören wir nirgends, daß die Walliser ein eignes Parlament beanspruchten oder dem englischen Nachbar feindlich wären. Der Walliser, auch in den engen Thälern, wo er seine Sprache frei von englischen Einflüssen bewahrt hat, ist stolz, ein Unterthan der Königin Viktoria zu sein, und gebeißt unter dem Schutze des britischen Löwen.

Anders die Iren. In Irland ist eine notleidende, unglückliche Bevölkerung, ohne Besitz, von oben bedrückt und von seinen Führern zur Gewaltthat dem Drucke gegenüber aufgereizt.

Wales wurde im Jahre 1282 von Edward I. erobert, nachdem Irland schon im Jahre 1169 unterworfen und im nördlichen Teile besiedelt worden war. Zeit genug haben die beiden Länder gehabt, sich an das Neue zu gewöhnen. Wenn sich Irland noch heute gegen England sträubt und die Treue

\*) Dieser Aufsatz wird unsern Lesern als Ergänzung zu dem im 19. Hefte: „Erinnerungen aus Irland“ willkommen sein.





# // *Die Grenzboten* – Fortsetzungsprojekt

## Aufgaben / Ansätze

- Deliverable 1: **OCR-Volltext-Nachbearbeitung**
  - Vergleich unterschiedlicher Softwaresysteme für Fraktur-OCR
  - Lückenidentifikation und -ergänzung auf der Ebene einzelner Seiten, Absätze oder Zeilen
    - ... Liste historischer Wortformen
  - Nachkorrektur von OCR-Volltext über Crowdsourcing
  - OCR-Volltext-Nachbearbeitung
  - Bewertung von Tools aus dem Projekt IMPACT – Improving Access to Text
- Deliverable 2: Integration des **Grenzboten-Korpus in CLARIN-D / DTAQ**
  - „Zoning“: Erfassung der textformalen Struktur der gedruckten Vorlage
  - Integration des Grenzboten-Korpus in CLARIN-D / DTA
- Deliverable 3: Erstellung eines **Erfahrungsberichts**

# Vielen Dank für die Aufmerksamkeit!

## Mitwirkende am Projekt „Die Grenzboten“

- Jan Paul Bultmann
- Faraz Ehssam
- Albert Gelter
- Dr. Maria Hermes
- Saskia Horn
- Dr. Manfred Nölte
- Magdalena Turkiewicz



## Mitwirkende am Fortsetzungsprojekt

- Dr. Martin Blenkle
- Dr. Maria Hermes
- Dr. Manfred Nölte
- Christian Thomas
- Gerrit Bruns
- Jan Paul Bultmann
- Maik Schünemann
- Kay-Michael Würzner