

Digitale Rätoromanische Chrestomathie

Eine réécriture communautaire

Jürgen Rolshoven, Florentin Lutz, Claes Neuefeind, Fabian Steeg

Sprachliche Informationsverarbeitung, Universität zu Köln
Universitätsbibliothek Köln (USB)

09. Dezember 2010



Réécriture communautaire

Das DRC-Projekt
Ähnliche Projekte

Methodisches Vorgehen

Digitalisate
OCR
Melioration

Umsetzung & Potentiale

Umsetzung
Potentiale



Réécriture communautaire

Das DRC-Projekt

Ähnliche Projekte

Methodisches Vorgehen

Digitalisate

OCR

Melioration

Umsetzung & Potentiale

Umsetzung

Potentiale



Ziele, Partner, Finanzierung

- ▶ Ziele:
 - ▶ Digitalisierung der Rätoromanischen Chrestomathie
 - ▶ Open Access zu Daten & Werkzeugen
- ▶ Projektpartner:
 - ▶ Köln: Spinfo & USB
 - ▶ Schweiz: Sprachgemeinschaft und Interessierte
- ▶ Finanzierung:
 - ▶ Köln: DFG
 - ▶ Schweiz: Legat Cadonau, IKG, Kanton GR
- ▶ Unterstützung:
 - ▶ Schweiz: SR



Réécriture communautaire

► réécriture

- ▶ la correction / l'amélioration
- ▶ Il faut noter l'importante différence sémantique qu'il y a entre amélioration et correction.
- ▶ l'intertextualité
- ▶ la translation / la traduction
- ▶ réécriture (informatique)

► communautaire

- ▶ Sprachgemeinschaft & Interessierte
- ▶ Computerlinguisten & Linguisten ... & Kulturwissenschaftler



Decurtins und die Chrestomathie

- ▶ Ca. 8000 Seiten aus 4 Jhd.
- ▶ Variantenreichtum
- ▶ Rezeption: Nicht alle Bündnerromanen lesen RF
- ▶ Nachdruck Octopus
- ▶ Register



Réécriture communautaire

Das DRC-Projekt

Ähnliche Projekte

Methodisches Vorgehen

Digitalisate

OCR

Melioration

Umsetzung & Potentiale

Umsetzung

Potentiale



- ▶ NZZ-Archiv
- ▶ ANDP
- ▶ Kleinsprachen
- ▶ Google



NZZ-Archiv

- ▶ 225 Jahre
 - ▶ 2 Mill. Seiten
 - ▶ PDF: Metadaten, Volltext
 - ▶ ReadOnly



Australian Newspapers Digitisation Program

The screenshot shows the Trove newspaper digitization program interface. At the top, there's a search bar and navigation links for 'Newspaper Home', 'About', 'Browse', and 'Help'. Below that, the main header reads 'Trove Digitised newspapers and more' with a sub-header 'Australian Newspapers is now part of Trove'. On the left, there are search fields for 'Find an article' and 'Find an issue', and dropdown menus for 'by Title', 'by Date' (showing months from 1803 to 2012), and 'by State' (listing states: NT, WA, SA, ACT, QLD, NSW, VIC). A map of Australia highlights the states. On the right, a thumbnail of a newspaper page from 'THE COURIER-MAIL (BRISBANE, QLD, 1948-...)' is shown, dated 'THURSDAY 2 DECEMBER 1948'. The page features a large headline 'FOUNDATION FOR MEMORIAL' with two black and white photographs of people. Below the page, there's a 'Recent comments' section and a 'Recent tags' section. At the bottom, a developer's view shows the page source code.

trove.nla.gov.au/newspaper



Kleinsprachen



vll.ladintal.it



Google

DÄNIELI BONIFACL

Catechibiu. Uadtil lit igl Meife<<, tru bkau Irdig Brea. IM 1.
(Abgedrnoki Bom>nia IX, 260 lt & Ulrich, Bh&torom. Texte I.)

[f. a'] ALS BEINADATCHEVS, NIEBEL8, STATTEVEL8, PRVS,
SABIGS ET HVNDREFELS S. Vugaa & Signurs d'üna intiera Drei- 5
chüra & Commün da Fur/tenotv, als meas oravqunt Hundrevels & chears
Singnurs & buns amigs, falüd da Deu, pä/ch & beinej/er tras nojs
Signer lefum Chrißum.

Slaunt cha la fanghia fcritttra^ Hundrevels Signurs ^ da per tntt igl
Mund eintin da tuUsa forts linguaghs ees rafad' ora & meffa per fcritt^ da iQ
tal fort & afchi clsBr, cha bigchia nam> ils Doctnrs & Muffa^ s da quella,
la cognufcan^ mh era ^ ute commün pievel & infaunts tras las fchkolas &
CatecbifemSi quegl ees^ curtas formas & compigliameints da tutts principals
punctgs della Chriftianevla Gretta & Religiun vegnan multaas & intragli-
daas^ ch'els quella ear cognufchan & tier üna veera cretta vegnan tratgs 15

...

tr>r Hj, igl quäl ees k Deu igl plij graund bein plafchgr. Sur quegl fc6

Bitoronuuitehe Ohreatomftthle. ^

Digitized by LjOOQIC



Vergleich der Ansätze

	Google	NZZ	ANDP	CLL	DRC
OCR	+	+	+	+	+
Quantität	++	+	+	--	--
Korrektur	--	-	+	+ -	+
Anreicherung	--	-	-	--	+

DRC fokussiert Partizipation:
 Stai si, conquista, romontsch, tiu vegl lungatg

Massen- und smarte Medien



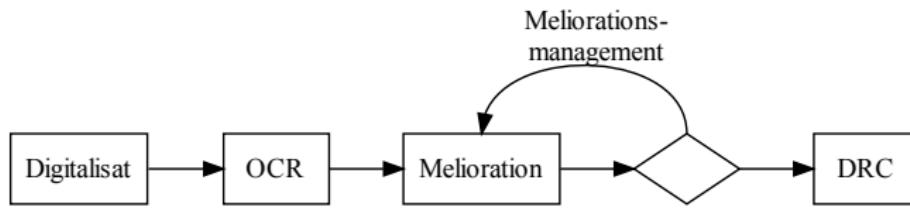
Réécriture communautaire
Das DRC-Projekt
Ähnliche Projekte

Methodisches Vorgehen
Digitalisate
OCR
Melioration

Umsetzung & Potentiale
Umsetzung
Potentiale



DRC-Workflow



Réécriture communautaire

Das DRC-Projekt

Ähnliche Projekte

Methodisches Vorgehen

Digitalisate

OCR

Melioration

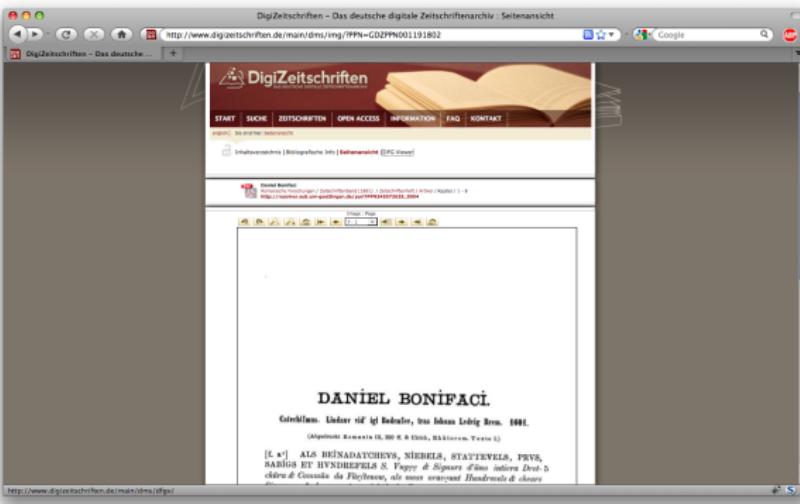
Umsetzung & Potentiale

Umsetzung

Potentiale



Quelle: Digizeitschriften



www.digizeitschriften.de



Scanner und ScanRobot



Réécriture communautaire

Das DRC-Projekt

Ähnliche Projekte

Methodisches Vorgehen

Digitalisate

OCR

Melioration

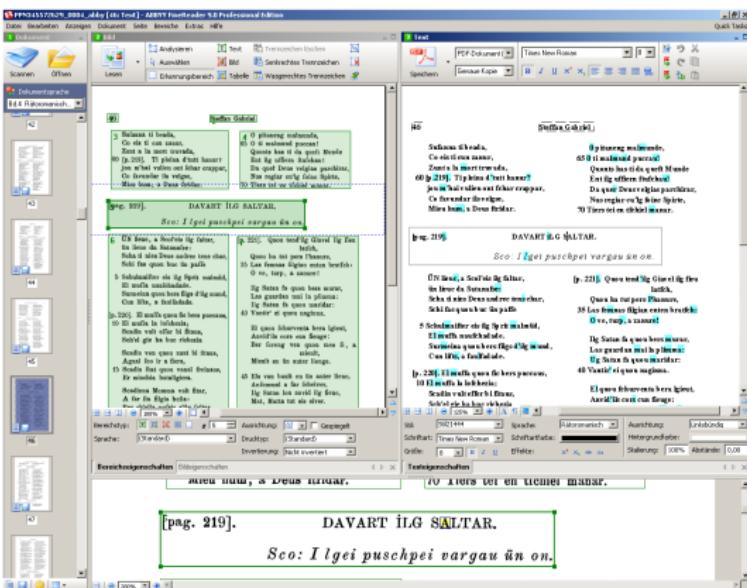
Umsetzung & Potentiale

Umsetzung

Potentiale



Abbyy Finereader



OCR-Workflow

- ▶ Eingabe: Digitalisate
- ▶ Mustertraining, Feinjustierung der Rahmen
- ▶ Ausgabe: PDF (Image, Text eingebettet)
→ Bild, Text und Metadaten (Wortkoordinaten etc.) werden in einem Meliorationswerkzeug zusammengeführt



Réécriture communautaire

Das DRC-Projekt

Ähnliche Projekte

Methodisches Vorgehen

Digitalisate

OCR

Melioration

Umsetzung & Potentiale

Umsetzung

Potentiale



Der DRC-Editor

The screenshot shows the DRC-Editor application interface. At the top, there's a search bar with 'bonifaci' and a dropdown menu for 'Volume Page Text'. Below it is a list of 6 hits, each with a preview and a timestamp (e.g., Thu Oct 14 16:32:45 CET 2010). The main window displays a document titled 'DANIEL BONIFACI' with the subtitle 'Catechismus Lindau vid' igl Bodenfer, tras Iohann Lydrig Brem. 1601.' The text is in Latin and includes some German. The bottom right of the main window shows a small watermark of the University of Cologne seal. To the right, there's a 'Page Comments' section with a table for 'Comment', 'Author', and 'Date'. Below the main window is a 'Check' section with a large green header 'DANIEL BONIFACI.' and a smaller subtitle 'chifimus. Lindau vid' igl Bodenfer, tras Iohann Lydrig Brem. 1601.' A note below says '(Abgedruckt Romania IX, 250 ff. & Ulrich, Rhoteron. Texte I.) [f. ar]' and 'ALIS BEINADATCHEVS, NIEBELS, STATTIEVELS, PRVS, SABOGS ET HVNDRFELS S. Vigan & Signers dura intenc Dext 5 chira & Ceteris da FirsSnows, als mess ova'nt Hundrevels & chears Signers & bon amige, Solid da Dea, p'ch & belesse tras nos! Signer iomM Christum.' The bottom left contains a 'Special Characters' keyboard and a 'Word History' table.

Form	Author	Date	Votes	Upvotes	Downvotes	Revert
BONIFACI	Ronenit Lutz from Switzerland (itz_01)	Thu Dec 02 21:25:21 CET 2010	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
BONIFACI	Jürgen Rohrbach from Germany (jrohrbach)	Thu Dec 02 21:22:40 CET 2010	-1	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
BONIFACI	Ulrich Steeg from Cologne, Germany (Ulrich_01)	Thu Dec 02 21:21:47 CET 2010	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
BONIFACI	Claes Neuendorf from Cologne (claus_1)	Thu Dec 02 21:17:14 CET 2010	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
BONIFACI		Thu Oct 14 16:35:47 CET 2010	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Beta-Version (Standalone)



Korrektur

Screenshot of a digital editing interface showing search results for "bonifaci".

Search Results:

- 6 hits for: bonifaci
- Volume Page Text
- Tags: 1601
- Comments: Erste Seite der RC (Bots: Downvote nötig für letzte Korrektur) Class Neufeld from Cologne Wed Dec 01
- Current page: volume 1, page 1, tagged as: 1601
- Add tag: [button]

Text Preview:

DANIEL BONIFACI. Catechisma. Lindau vid' igl Bodenfee, tra Johann Lwig Brem. 1601.

(Algebraic Romania IX, 250 ff. & Ulrich, Rätoroman. Texte I.) [f. ar.]
ALIS BEINADATCHEVS, NIEBELS, STATTIEVELS, PRVS, SABIGS
ET HVNDREFELS S. Vugia &^a Signars d'ite intera Diet- 5 chira &
Cossini da Pfisnern, als mose oev'ont Handwrels & chears Signars
& bons amig, Salld da Des, pach & heiness tra nos! Signar iemM
Christen.

Sicut cha la Sauglia Scritina, Handwrels Signars, da per tutt igl
Mund storia di tutte forte lingue om rafal' ora & metta per Scrit-
da 10 tal Sort & afchi clair, cha binchia namec li Ductos & Manus
Christen.

Entry Details: Entry 'BONIFACI.' is locked (vote down to unlock)

Special Characters:

À	Á	Â	Ã	À	Á	Â	Ã	À	Á	Â	Ã
à	á	â	ã	à	á	â	ã	à	á	â	ã
È	É	Ê	Ë	È	É	Ê	Ë	È	É	Ê	Ë
è	é	ê	ë	è	é	ê	ë	è	é	ê	ë

History:

Form	Author	Date	Votes	Upvote	Downvote	Revert
BONIFACI.	Roman Lutz from Switzerland (luz, 0)	Thu Dec 01 21:25:21 CET 2010	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
BONIFACI.	Ulrich Reichenauer from Cologne, Germany (Ulrich, 0)	Thu Dec 02 21:22:48 CET 2010	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
BONIFACI.	Felix Steeg from Cologne, Germany (Fotreg, 1)	Thu Dec 02 21:19:19 CET 2010	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
BONIFACI.	Class Neufeld from Cologne (classen, 1)	Thu Dec 02 21:17:14 CET 2010	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
BONIFACI.		Thu Oct 14 16:35:47 CET 2010	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Grammaire des fautes

Automatisierung: Edit suggestions (Lexika + Edit-Distance)



Korrektur

035 - 008 42 i ftdij. abbeverare le beftie. A. A. avunda.
 036 - 026 75 Per viver et morir beada. 175 Pro temp regueva I
 011 - 021 15 ton de queschier E ton de surportar. 5 La mumma 1'
 012 - 016 59 las giias stouvan strasuner, 25 E Par nus tingner
 030 - 019 41 enten Jesum Christum, udit il plaid da Deus,
 037 - 029 17 Ilg Mastral da Criminal dûrond feis offici non s def
 012 - 027 11 Pg afchaid, fchi dis el: Elg es cumplieo: & hauiland
 012 - 029 40 èg liqfutichad. Deis ils upewris fplittimel,
 017 - 066 15 dal utfché, è tuot las chantaditras veguen à gnir
 017 - 047 35 ilg fo3 dalg infierme. 112. Cun tuot avautant Deis ns'
 017 - 055 37 E lur fadifs ils han fquifchads, & fut lur man fun eus
 024 - 007 14 fch' els fuorien informâts li nobel ftdi del Baur non
 017 - 013 40 giagliard in la mtidaeda cun la queila ttu vaint
 037 - 032 5 lugiu Laudama dûrond feis officij non defts tant
 035 - 008 33 Romancio di Surfvela Italiano A. A (Propofitione)
 i ftdij. abbeverare le beftie. A. A. avunda.
Abadeffa. Avat. bandunar. darsi li ftdi. boventâ. Romancio di
Abadefsa Adelheit, nativa da Novena, Prudainta et pia confeſsa
abadessa, Ha in faseler tgau. Ella scogna avon messa E sco ch'
abadiennt, jal fat CunPieu fun alg mendar TimP, schi
abadiit à quell, miliguret vofas vittas, vas parrieit da 40 morir, vivit
abafar de gradu ad ir cun (s h) raanadûras da dor punt Martina
abaffò Pg chio fchi def 10 el fu Pg fpieri. Et Ps ludeaus par ch' elg
abaffa Ig adutzad. Parché chia tti inuidachis, Meis Deis chi num m'
abaffadas. 7. Et eir il craifta vain à tmair das louhs sti aut, ed
abaffain Da eottr, chi ais grond tempe, 780 Et nofs puehiads tuots
abaffats. 43. Bleras votas ils ha el deliberats; ma [p. 309] eus il
abaffeSen quel uScheea pro tuottas natiuns las plu iluminâdas
abaffo. Perche fco Pg oç run l'go feie vain appriuo, vfchea li fieud
abafsar a far li fuorium flegucia or da nofs pajais tant dominant
abafanza abbadezza, abbate, abbandonare, abbandonar li ftdii

KeywordInContext: Types zur Lesefehlersuche



Anreicherung

Screenshot of a digital edition interface showing the text of Daniel Bonifacius' *Carthorum Lindau vid' igl Bodenfee*. The interface includes a sidebar for navigating the volume and page, a main text area with a checkmark for 'Daniel Bonifaci', and a right panel for 'Page Comments' and 'Word History'.

Page Comments:

Comment	Author	Date
Erste Seite der RC	Class Neuafind from Cologne	Wed Dec 01
[redacted]: Downvote nötig für letzte Korrektur	Florentin Lutz from Switzerland	Thu Dec 02

Word History:

Form	Author	Date	Votes	Updates	Downvotes	Revert
BONIFACI	Florentin Lutz from Switzerland (Eta, 0)	Thu Dec 02 21:25:21 CET 2010	0			
BONIFACI	Jürgen Röhlwien from Cologne, Germany (Pots, 0)	Thu Dec 02 21:22:45 CET 2010	-1	1	0	
BONIFACI	Class Neuafind from Cologne, Germany (Staten, 1)	Thu Dec 02 21:17:14 CET 2010	0	0	0	
BONIFACI	—	Thu Oct 14 16:35:47 CEST 2010	0	0	0	

Automatisierung: Lernen aus Nutzerverhalten (Tags)



Interaktion

The screenshot shows a digital edition interface with the following components:

- Search Results:** A sidebar on the left displays a list of search results for "bonifaci". It includes columns for Volume, Page, Text, Modified Date, and Tags. One result is highlighted: "DANIEL BONIFACI. Canticum. Lindau vid' igl Bodenfee, tras Iohann Lydrig Brem. 1601." The date is listed as "Thu Dec 02 21:25:21 CET 2010" and the tag is "1601".
- Page View:** The main area shows a manuscript page from "DANIEL BONIFACI. Canticum. Lindau vid' igl Bodenfee, tras Iohann Lydrig Brem. 1601." The text is in a Gothic script. Below the page, a note reads: "(Algedruckt Romania IX, 259 ff. & Ulrich, Rhatorom. Texte I) [f. 1r] ALS BRUDERFELVS. NIEBELS. STATTUEVELS. PRVS. SABICVS ET HVNDRIEFLS S. VIPSAS & Signari d'una intera Dint- 5 chris & Corinii de PfarSmow, al mens eou'ant Hendervels & chears Signari & bonz anisi, Salat da Des, pe'ch & beineser tra noi" Signari leuM Christum.
- Comments:** A floating window titled "Page Comments" shows a single comment: "Erste Seite der RC" by "Class Neuvald from Cologne Wed Dec 01".
- Word History:** A sidebar at the bottom right shows a history of the word "BONIFACI". It lists several entries with their authors, dates, and vote counts (Upvotes, Downvotes, Reverts). The most recent entry is by "Florentin Lutz from Switzerland" on "Thu Oct 14 16:35:47 CET 2010".

Automatisierung: Lernen aus Nutzerverhalten (Korrekturen)



Melioration

Melioration = Korrektur + Anreicherung + Interaktion

= Dialoge mit und über Texte



Réécriture communautaire

Das DRC-Projekt

Ähnliche Projekte

Methodisches Vorgehen

Digitalisate

OCR

Melioration

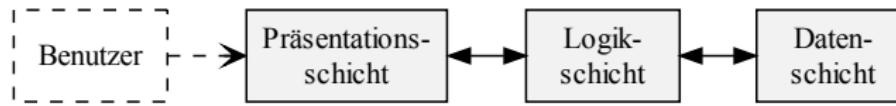
Umsetzung & Potentiale

Umsetzung

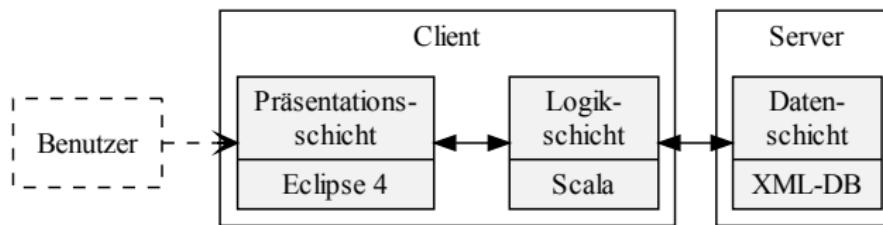
Potentiale



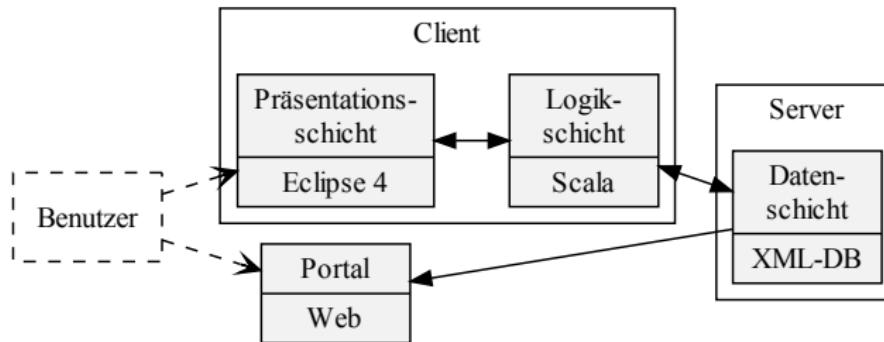
Architektur: Überblick



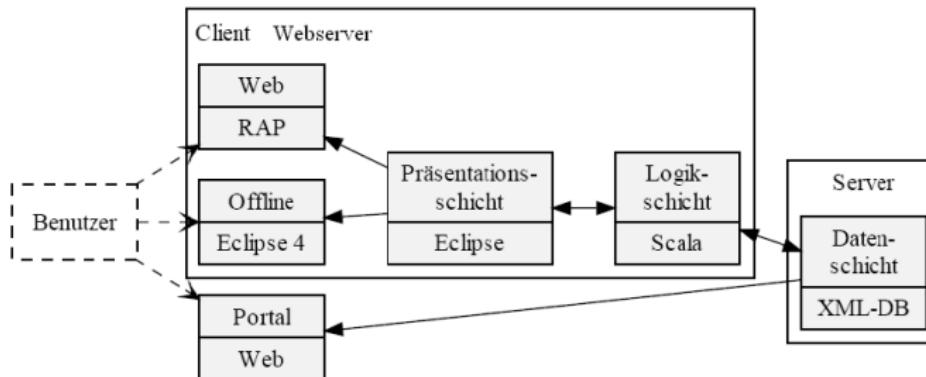
Architektur: Beta-Version



Architektur: Beta 2



Architektur: Planung



Réécriture communautaire

Das DRC-Projekt

Ähnliche Projekte

Methodisches Vorgehen

Digitalisate

OCR

Melioration

Umsetzung & Potentiale

Umsetzung

Potentiale



Linguistik: Chrestomathie und Lexikographie

1

012 - 003 38	A s1 tti fes, chel managia, fchi m1 pain our d' ed in uestig ed eUa ei en pauc temps morta della leu gauden. [p. 615] DaCielin Aungelperfacciar, Ina 273 Poesias..... ^69 Viola..... 269	malenconia. Iofeph 4. dich: Amich ilg fummi tieu wol dir aquaift: 280 malenconia. 113. La buoba ed il morder grond. Ina dunna giuvna erina malenconia. 75 A foing Francefc volent funar La fia melodia, Vet ftruch malenconia autuilla..... 269 Sentinella alla glina..... 270
011 - 040 20	leu bargir Entafuns il lac sper via Larmas spir fatg ton mitgiert cun il giuven, ch'el ei morts della fa attenzion. El passa tras ils vasts curtins, Els églis Landa fia Jesus Cristus Per tott l'etternitá. Schia di Guslyr 5. 50 eau nun se ünguota da que dir, La tia Perche mieu cour plajo eis huofs Per grand da sia aeted. Ach Dieu, mieu cour tres our. Per taunt da far menzien. Nha bain adiuua fitch temti 50 Piglå mia Tú per mai non plondscher plü, 35 Quia nun jtida ravaschia, Cur el, da plaschair, ais plain Mo dutsch tei teis char amur. Eu vegen davent da quia Cun grond eir il tschervé. 15 Hopsa jufifallerallera. 0 dolor, Barbla co Lai serius' as fo ?Volkslieder. Be per Giunfra Nuttinga Barbla co La serius' as fo Be per 78) Alla melanconia. (Suainter Lenau.) Sentimentel Eir el mort e sutterrà. Hopsa jufifallerallera. 0 dolor da quiaist muond! ?Gian Fadri Caderas. [p.78] Alla amih, Cha las chosas glajan, Scu chi voglian ir. 5 La de cheu ad in on vegnies ad els da quellas lunas in spounchia p lg wair, Ch' eau wygnia incuort granda	malenconia. D' ina mumma, che lamenta, Ch' ella seigi maltractada 235 malenconia. Mo la spusa saveva nuot, ch'el fuss morts. La sera dil di malenconia. Salida quel, salida tschel 180 Cun bialla cortesia, E stat lu malinconia. 30 Vofs cor sarà chiargia! If. 8rJ Oder d'allegraria. Sarà quel malinconia a m1 dalaist tū scuwnir! Lg Bab 6. [f. 2a] Meis filg saslungia malinconia, Psalm 129, Eau sum afflict et plain d' angouscha 22, 23 malinconia Sto be da schliper our, Nun so nun piglen via 5 Pal grand malinconia Uscheta non sa jeva ptli Il chiar fu tott or via. Seis modo da malinconia Usché voul il punt d' onur. Mia armada ais stin la plaza Et melanconia, 10 Il pù bel attrat containi! Eau li pù am poesia Cur la sort melanconia 25 Be per tai meis char cor! ?Volkslieder. 108. (Annalas melanconia, Amalá ma compagnia! Hopsa jufifallerallera. La furmia get melanconia, Cha non la dun la sia; 30 Ma uossa sun eau co. Maister melanconia, Cha non la dun la sia; 30 Ma uossa sun eau cò. Maister melanconia, Cun me chaminst per di, Cur mia staila bella brilla, E melanconia, Eu ha pers ma compagnia." 30 Hopsa jufifallerallera. Eu ha melanconia (Suainter Lenau.) Sentimentel melanconia, Cun me melanconia Zuond nu 'm da fer, A culöz per Dieu Nun sè 'l muond melanconias a defs da quei el fen o lu fufs ei eun fiu faigt, aber ascheia melinconia ad hauair. Sieu Bun Cuslier 20. 1435 chie chiofsa
030 - 030 46		
038 - 035 17		
008 - 049 55		
011 - 041 14		
008 - 047 49		
036 - 036 69		
012 - 005 9		
027 - 029 29		
027 - 028 23		
036 - 027 65		
027 - 025 24		
024 - 016 12		
027 - 016 31		
027 - 019 5		
027 - 027 2		
024 - 004 4		
024 - 028 5		
027 - 019 17		
024 - 028 3		
024 - 004 26		
030 - 048 32		
012 - 009 24		



Linguistik: Chrestomathie und Syntax

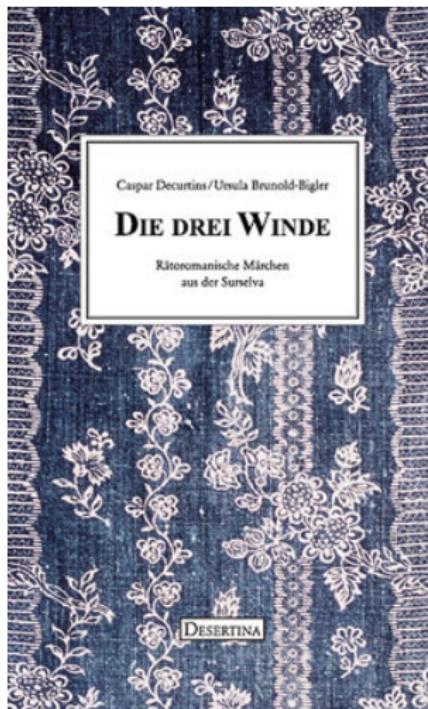
030 - 011 40 Mintgia cuitg aber feigi obligaus de render in giest quen als signrs giraus da quellas varts, *nua che* dei nuott auter puder vegni conpigliau en quen che agli segner farer per benedi lalp....

011 - 040 12 Las treis veglias, che dormevan tuttas treis en in letg e magliavan ord in cnp, mavan mintga notg gin en in grond tschaler, *nua ch'* ei era spir candeilas, che ardevan.

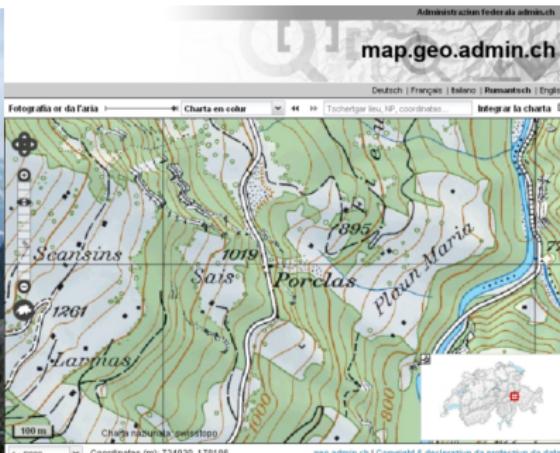
Denton tut che *selai* splenar, cura ch' ils sauns assistan als malsaus...
Orda nossa patri'alpina mai *selais* ti bandischar



Erweiterungen: Übersetzungen



Erweiterungen: Georeferenzierung & Crossmedialität



011 - 005 3 Volkslieder
 Äch co mi fa mal il cor,
 Che miu muronz ei jus da Porclas or.

Crossmedialität? (mp3)



Erweiterungen: Bündnerromanisch

- ▶ Gebrauchswörterbuch: Pledari grond
- ▶ Gemeinsame Textbasis u. gemeinsamer Zugriff DRC u. VLL
- ▶ Erweiterung Textbasis
 - ▶ vorhanden (OCR):
RC, Bifrun Gartnerausgabe, Tschantamaints, Puorger
 - ▶ vorhanden (Digitalisat): DRG
 - ▶ Literatur 16, 17, 18 Jh.
 - ▶ Zeitungen
 - ▶ Strukturierte Texte: Lexika
 - ▶ ??Trésor de la langue romanche??



Erweiterungen: Sprachen

DRC als Prototyp

Aromunisch

Kreolsprachen

Réécriture und Repatriation

Programmsystem dank RAP und Single Sourcing sowohl vernetzt
als auch Stand-Alone



Vernetzung

Engraziel fetg!

www.spinfo.uni-koeln.de/forschung-drc.html

rols@spinfo.uni-koeln.de

florentin.lutz@sunrise.ch

c.neuefeind@uni-koeln.de

fabian.steeg@uni-koeln.de

