

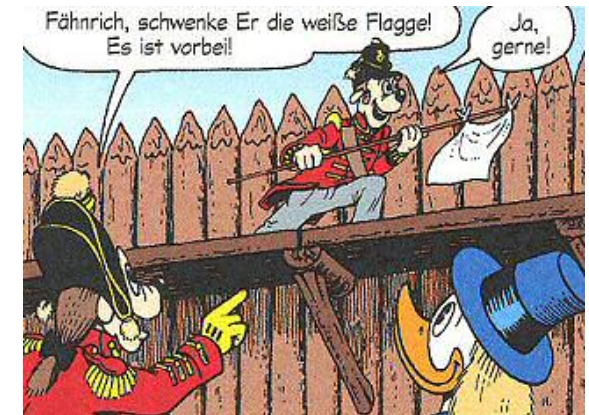
Überlegungen zu einem Betriebsmodell für etablierte wissenschaftliche Software

Thomas Schmidt (IDS Mannheim) &
Kai Wörner (Universität Hamburg)



Überlegungen zu einem Betriebsmodell für etablierte wissenschaftliche Software

- CFP: „Wer soll das bezahlen?“
- Arbeitstitel:
 - „(Re)kapitulation: Vom Ende und möglichen Neuanfang der Entwicklung einer wissenschaftlichen Software“
 - „Scheiße, es funktioniert“
- „Das angesprochene Problem ist ein sehr ernstes, denn zu viele gute wissenschaftliche Software verschwindet oder wird unbenutzbar, weil kein geeignetes Finanzierungsmodell gefunden werden konnte.“
[Kommentar eines Gutachters]
- Dieser Beitrag: keine Lösung, Fallstudie, erste Ideen



EXMARaLDA

- Erstellen, Verwalten, Auswerten, Publizieren audiovisueller Sprachkorpora
 - Gesprächskorpora
 - Spracherwerbsdaten
 - Soziolinguistische/Dialektologische Daten
- Transkription und Annotation (Partitur-Editor)
- Korpus-Management / Metadaten (Coma)
- Query/Analyse (EXAKT)

EXMARaLDA EXAKT 1.1

DEMO-KORPUS (434 results)

#	Communication	Speaker	Left Context	Match	Right Context	Remarks	Beruf[S]
18	Rudi Völler: Wutausbruch	RV	I Wir ham heute .. Und da hat	der	Sebastian Kehl .. Wir ham + genei		Teamchef
19	Rudi Völler: Wutausbruch	RV	.. Wir sind Vize Weltmeister ..	das	muss n bisschen mehr kommen +	verz. enclitd.	Teamchef
20	Rudi Völler: Wutausbruch	RV	en .. Aber diesen Scheiß ..	der	da immer tabellar wird + Da so	Tea KWIC concordance	
21	Rudi Völler: Wutausbruch	RV	irklich mal Gedanken machen, ob	das	/ + ob wir in der Zukunft so wei	wrong case	Teamchef
22	Rudi Völler: Wutausbruch	RV	sien machen, ob das + ob wir in	der	Zukunft so weitermachen können?		Teamchef
23	Rudi Völler: Wutausbruch	RV	ies runter zu ziehen? - Das ist	das	Alterstetel + Und ich lass mir		Teamchef
24	Rudi Völler: Wutausbruch	RV	Alterstetel? - Und ich lass mir	das	nicht mehr so lange gefalten. D		Teamchef
25	Rudi Völler: Wutausbruch	RV	.. Ach, wieso müssen wir denn?	das	Mannschaften klar bezeichnen?		Teamchef
26	Rudi Völler: Wutausbruch	RV	ich hab doch ich hab doch	die	Mannschaft .. Neint ich hab doch		Teamchef
27	Rudi Völler: Wutausbruch	RV	Mannschaft .. Neint ich hab doch	die	Mannschaft ich hab doch die Ma		Teamchef
28	Rudi Völler: Wutausbruch	RV	ich hab doch ich hab doch	die	ich hab doch ich hab doch	repair sequence	Teamchef
29	Rudi Völler: Wutausbruch	RV	isch. Am Mittwoch werden nur	das	Leute spielen .. die sich wir		Teamchef
30	Rudi Völler: Wutausbruch	RV	iden nur die Leute spielen ..	das	sich wirklich hundertprozentig		Teamchef
31	Rudi Völler: Wutausbruch	RV	st doch tuhst! Der Gürtel! Was	das	früher für n Scheiß gespielt ha		Teamchef

Remarks: repair sequence
Beruf[S]: Teamchef der Fußballnationalmannschaft seit 2000

Partitur

RV [v] ich hab doch die Mannschaft kritisiert. Is doch ganz klar. S war heute nich in Ordnung. • S war zu wenig Laufbereitschaft. Ar

RV [en] wasn't playing well. That's pretty obvious. That was not okay today. • There wasn't enough willingness to run. W

RV [nv] lehnt sich wieder zurück Gestik im Folgenden weniger aggressiv und hektisch, d

WH [ki]

EXMARaLDA Coma | EXMARaLDA, DemoKorpus.coma | Corpus: DEMO-KORPUS

Korpus | Daten | Korpus-Korb (0) | Einstellungen

Filter: Alle Filter löschen | 18 Kommunikationen

Kommunikation

S	Name	Transkription erstellt in
	Rudi Völler: Wutausbruch	Januar 2004
	Hilge Schneider: Arbeit...	?
	Huber Fichte: Interview	November 2003
	Hilge Schneider: Tropf...	?
	Studio Braun: English T...	August 2008
	Deutsche Bahn: Rossau	August 2008
	Studio Braun: Forum Wa...	Oktober 2008
	Sigalène Royal: Débat	November 2008
	Fernando Savater: Her...	November 2008
	Paul McCartney: Interview	Dezember 2001
	Seif Iasi: Experimentauk...	November 2008
	near story	April 2009
	Henry Fröhlich: My Theory	Januar 2009
	Hart aber fair	Juli 2009
	Anne Will: Halbes Wahr...	Dezember 2009
	Goepferle: Cello	Januar 2010
	Nyuyen Ngoc Ngan	Januar 2010
	Tella-Telenor	September 2000

Suchen

Filter: Alle Filter löschen | 45 Personen

Communication (Communication) 2

Description (Communication) 2

Aufnahmedatum: 07 09 2003

Gesprächstyp: Fernsehinterview

Quelle: Sportschau (ARD) am 06.09.2003. Aufgenommen und bereitgestellt von Dr. Wilfried Schütte (Institut für Deutsche Sprache, Mannheim)

Transkribent: Annette Schneider

Transkription erstellt in: Januar 2004

Vorgeschichte: Das Fußball-EM-Qualifikationsspiels gegen Island endete 0:0. Vor der Live-Schaltung nach Reykjavik haben die Moderatoren der Sportsendung, Gerhard Delling und Günter Netzer, die nach auf einem Monitor im Studio zu sehen sind, das Spiel bzw. die Leistung der deutschen Nationalmannschaft bewertet.

folder: Rudi

project-name: EXMARaLDA DemoKorpus

HAT (vereinfacht), für Nonverbales projektspezifische: Mimik und Gestik der Sprecher werden nur analysiereweis angedeutet. Abkürzungen: LA= linker Arm, RA= rechter Arm, LH= linke Hand, RH= rechte Hand, KO= Kopf, OK= Oberlippe

transcription-convention: HAT (vereinfacht), für Nonverbales projektspezifische: Mimik und Gestik der Sprecher werden nur analysiereweis angedeutet. Abkürzungen: LA= linker Arm, RA= rechter Arm, LH= linke Hand, RH= rechte Hand, KO= Kopf, OK= Oberlippe

transcription-name: Rudi Völler Wutausbruch

Location 2

City: Reykjavik

Country: Island

PeriodStart: 06.09.2003 21:30:00

Description (Location) 2

Lokal: ARD Fernsehstudio

Alle Filter löschen | 45 Personen

Sigle: Beruf

ERW: Verfahrenstechniker

Fichte: Schriftsteller

KL: Arbeitslos

KV: Teamchef der Fußballnationalmannschaft seit 2000

WH: Sportjournalist seit 1970

UB: "Wirtschaftler in Deermwohnheim", Verdienst 1000-1...

AMT: Fachangestellter für Arbeitsförderung

BS: Programmierer

TT: Dolmetscher

JM: Hauptbuchhalter

GS: Maurer

PF: ?

J: Steuerfachgehilf

BP: Kfz-Mechaniker

SR: Politiker

HS: Staatspräsident

PP: TV-Moderator

AC: TV-Moderatorin

FS: Schriftsteller

LR: Radiomoderator

INT: Radiomoderator

PMC: Musiker

MOD: Experte für soziale Fragen

AW: Rentnerin

X: ???

Y: ???

PR: TV-Person

ELK: Theoretik

FP: Fernsehmoderator

KP: CDU-Generalsekretär

PK: Hausfrau

FB: Rentnerin

EK: Geschäftsführerin

HB: ???

AW: Fernsehmoderatorin

EXMARaLDA Partitur-Editor 1.4 [T:\TP-ZPIDATHEXMARaLDA_DemoKorpus\PearStory\PearStory.xml]

File Edit View Transcription Tier Event Timeline Format Segmentation Help

And then you see ehm a maan in maybe

00:00 00:01 00:02 00:03 00:04 00:05 00:06 00:07 00:08 00:09 00:10 00:11 00:12

0 [00:0] 1 [01:9] 2 [01:9] 3 [03:2] 4 5 [09:2] 6 [10:4]

X [v] So it starts out with: A louder

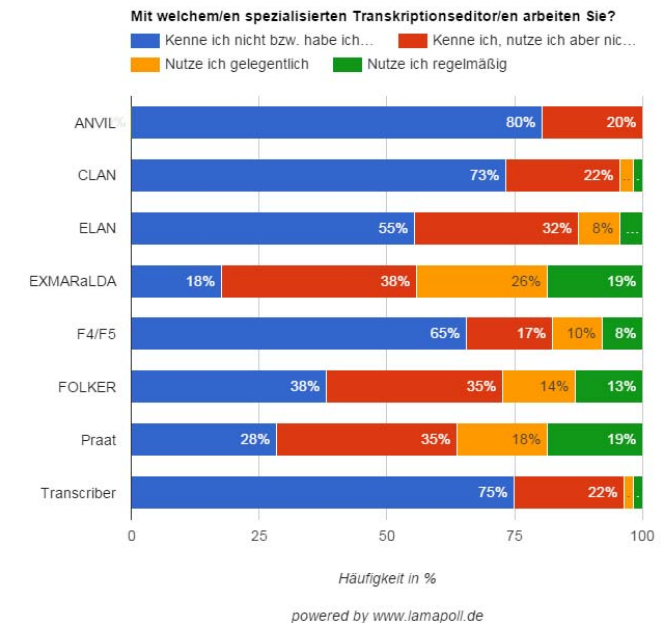
X [nv] IHA on IHN, IHA on ISH IHA up and to the right IHA stays up IHA back down IHN, moves IHA on ISH IHA to the right IHA up, rH

X [nv] emphasizes the crow HE nods once

Y [v] approximately fifty

Verbreitung / Verwendung


- 500-1000 Downloads pro Monat, seit vielen Jahren
- 100-200 Support-Anfragen jährlich über CLARIN-Helpdesk
- Nutzerstudie „Mündliche Korpora“: 19% nutzen EXMARaLDA regelmäßig, 26% gelegentlich, nur 18% „kenne ich nicht / habe ich noch nie genutzt“
- Basis mehrerer großer Korpora
- Regelmäßige Anfragen für Schulungen, von vielen verschiedenen deutschen Universitäten, aus 15 verschiedenen Ländern
- Als Best-Practice empfohlen in DFG-Handreichungen



Wer hat es bisher bezahlt?

Finanzierung / Projektzusammenhänge

- 2000-2011: SFB 538 Mehrsprachigkeit
- ab 2011: HZSK Hamburg (CLARIN/LiS) / IDS Mannheim (AGD)
- Mischfinanzierung: Universität Hamburg, Qualitätsoffensive der Stadt Hamburg, DFG-SFB, DFG-LiS, BMBF-CLARIN, Leibniz
- Software-Entwicklung immer als Mittel zum Zweck, ein Arbeitspaket unter mehreren
 - Computergestützte Erfassungs- und Analysemethoden multilingualer Daten (SFB)
 - Etablierung eines Schwerpunkts Mehrsprachigkeit und gesprochene Sprache am HZSK (LiS)
 - Archiv für Gesprochenes Deutsch (IDS)
 - Eine web- und zentrenbasierte Forschungsinfrastruktur für die Geistes- und Sozialwissenschaften (CLARIN)

Sonderforschungsbereich
Mehrsprachigkeit 

 INSTITUT FÜR
DEUTSCHE SPRACHE

 Bundesministerium
für Bildung
und Forschung

 DFG

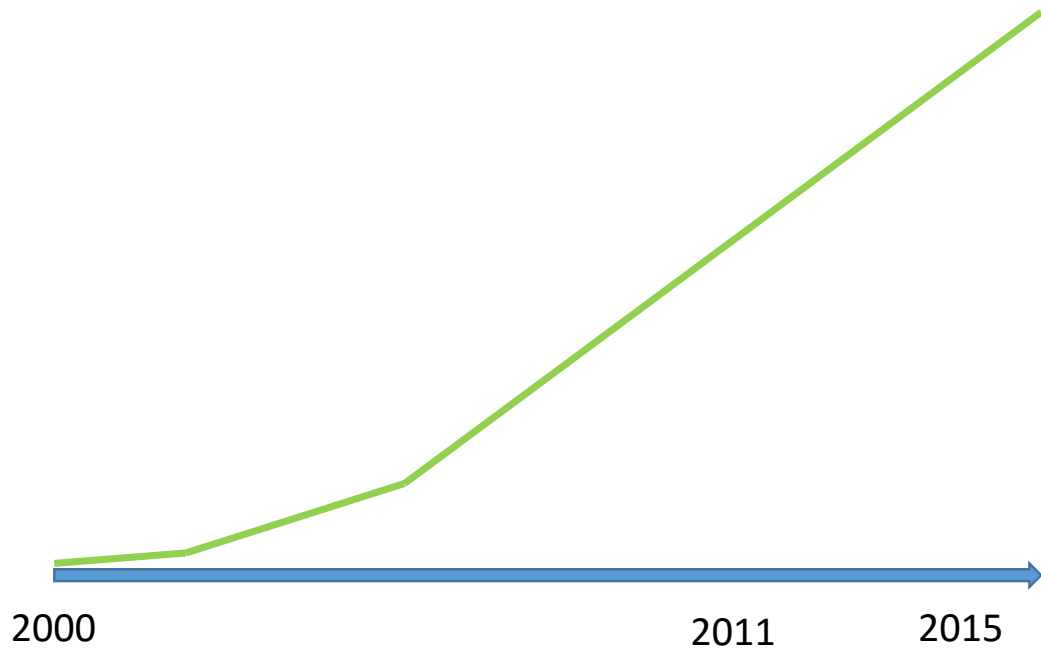
 hzsk

 CLARIN-D

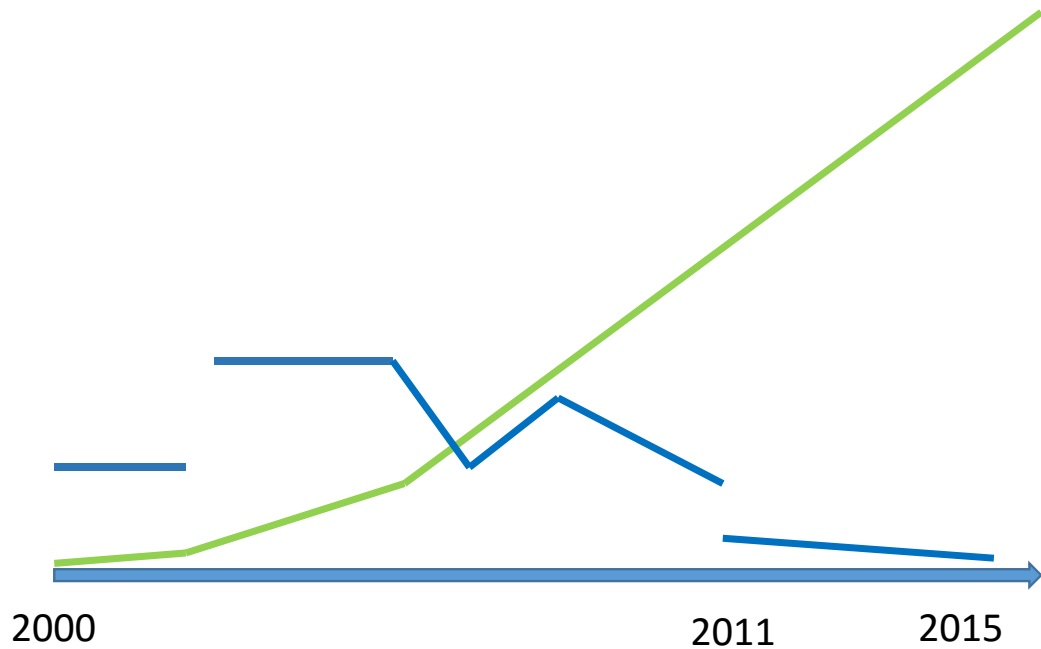
Nutzung/Entwicklung

Anzahl der Nutzer

- 2 zu Beginn
- Hunderte nach ca. 5 Jahren (Basisfunktionalität)
- Tausende aktuell



Nutzung/Entwicklung



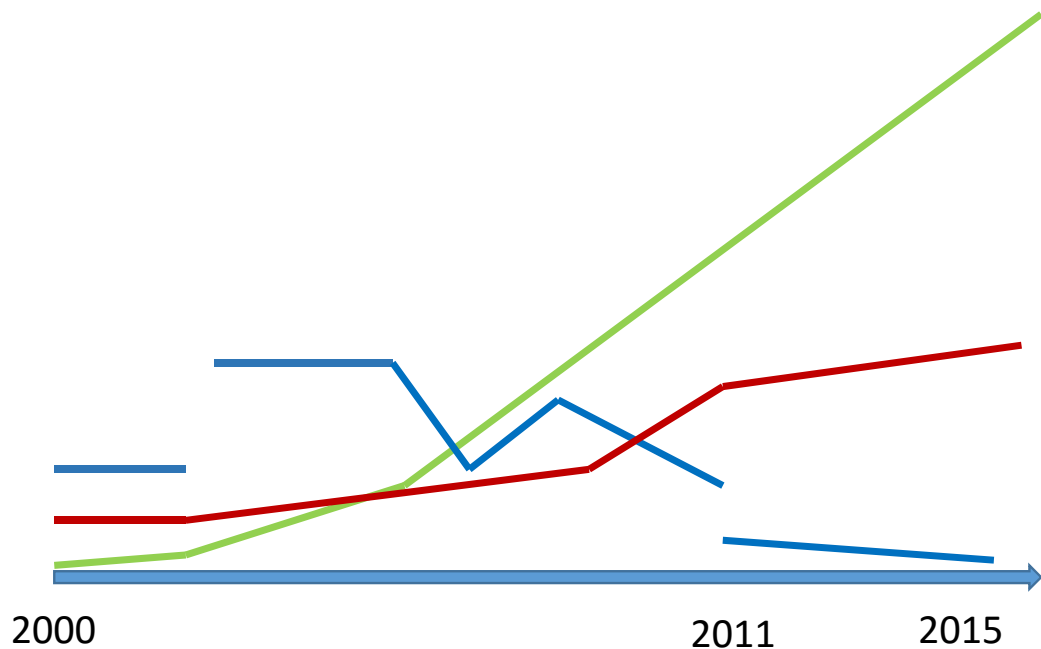
Anzahl der Nutzer

- 2 zu Beginn
- Hunderte nach ca. 5 Jahren (Basisfunktionalität)
- Tausende aktuell

Wöchentlich investierte Entwicklungszeit (inkl. Maintenance)

- ca. 30h zu Beginn (initiales Design, Prototypen, 1 Person)
- ca. 60h 2003-2005 (Ausbau, 2 Personen)
- Danach weniger, schwankend, je nach aktueller Priorität im Projekt (Datenkuration im SFB, bis zu 4 Personen)
- Seit SFB-Ende deutlich weniger, aktuell ca. 1h
- Regelmäßige Updates bis 2011, seitdem nur noch sporadisch/"inoffiziell"/auf Zuruf

Nutzung/Entwicklung



Anzahl der Nutzer

- 2 zu Beginn
- Hunderte nach ca. 5 Jahren (Basisfunktionalität)
- Tausende aktuell

Wöchentlich investierte Entwicklungszeit (inkl. Maintenance)

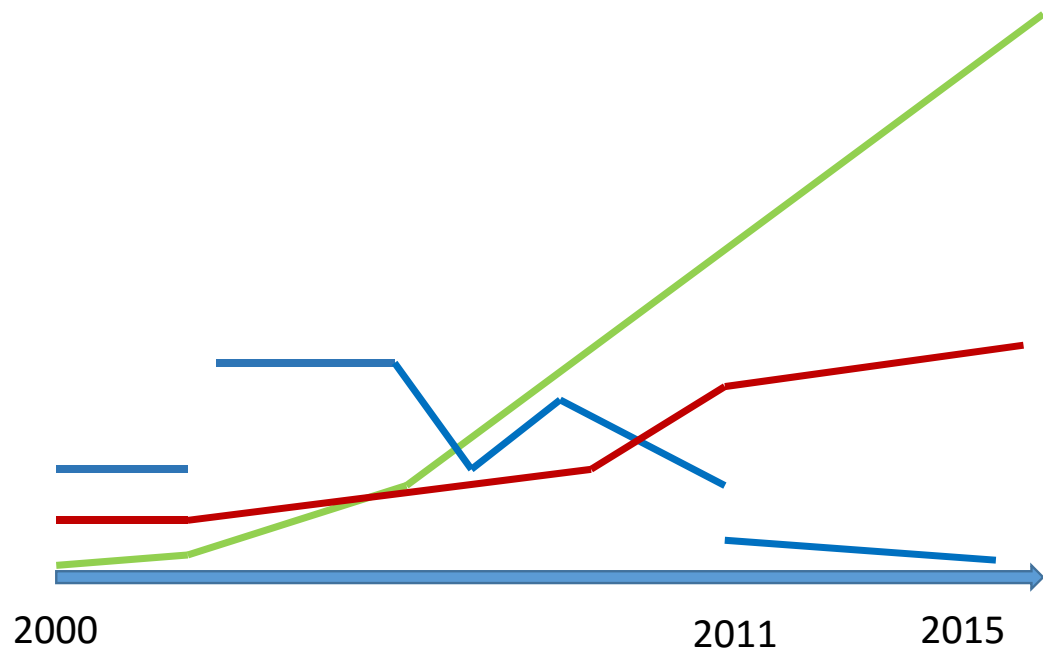
- ca. 30h zu Beginn (initiales Design, Prototypen, 1 Person)
- ca. 60h 2003-2005 (Ausbau, 2 Personen)
- Danach weniger, schwankend, je nach aktueller Priorität im Projekt (Datenkuration im SFB, bis zu 4 Personen)
- Seit SFB-Ende deutlich weniger, aktuell ca. 1h
- Regelmäßige Updates bis 2011, seitdem nur noch sporadisch/"inoffiziell"/auf Zuruf

Supportbedarf/-aufwand

- Individuelle Betreuung zu Beginn
- Wachsende Zahl an E-Mail Support-Anfragen
- Schulungen seit 2008
- CLARIN-Helpdesk seit 2011

Nutzung/Entwicklung

- ❖ Wachsende Nutzerzahlen
- ❖ Anhaltender Supportbedarf
- ❖ Schwindende Investition in Entwicklung/Maintenance



Aktuelle Aufgaben

Maintenance

- Anpassung an neue OS
- Unterstützung neuer Videoformate
- Touchscreens
- Umstieg SourceForge > Git
- Bugs

...

Ausbau (klein)

- ✓ neue Importfilter
- ✓ Usability-Adaptionen

...

Ausbau (groß)

- Anbindung an CLARIN-Webservices (WebLicht)
- Kompatibilität mit ISO/TEI Standard
- Neue Annotationsmethoden
- Neue Datentypen
- Übertragung in webbasierte Architekturen

...

Dokumentation / Support / „Marketing“

- Schulungen / Email-Anfragen
- Nachführen von Manuals etc.
- Internationalisierung
- Präsentation auf Konferenzen / Fachmessen

...

- Kleinere Feature Requests können noch relativ problemlos umgesetzt werden
- Innovative Entwicklungen nur noch schleppend oder gar nicht
- Erhebliche Verspätung bei kritischen Anpassungen
- Dokumentation hinkt immer weiter hinterher
- „Service“-Qualität lässt nach

Weiter so?

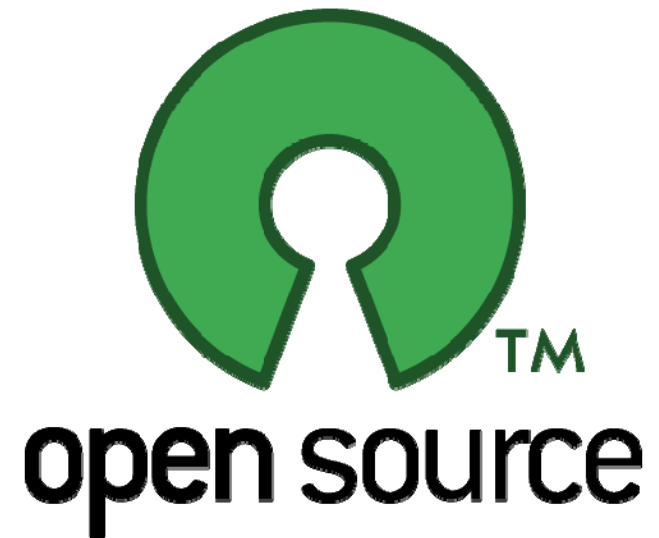
- Zunehmende Instabilität
 - Erhöhter Support-Bedarf bei
 - Ausbleiben von Anpassungen an Neuerungen
 - Zu langsamer Nachführung der Dokumentation
 - Nachlassende „Kunden“-Zufriedenheit
- ➔ Kapitulation: Abandonware (vgl. Transcriber)

- Bedeutung für universitäre Ausbildung (Curricula), laufende Forschung
- Bedeutung für einschlägige Datenzentren, Infrastrukturen, Best Practices, Standardisierungsprozesse
- Situation ähnlich für „Konkurrenten“ (ELAN)



Open Source?

- EXMARaLDA ist OpenSource
- ca. 10 punktuelle Beiträge innerhalb von 10 Jahren
- Organisch gewachsene Code-Basis
 - JDK 1.2 / kein DOM-Parser für XML in 2000
 - keine Refaktorisierung in der Zwischenzeit
- Wer sollte beitragen???

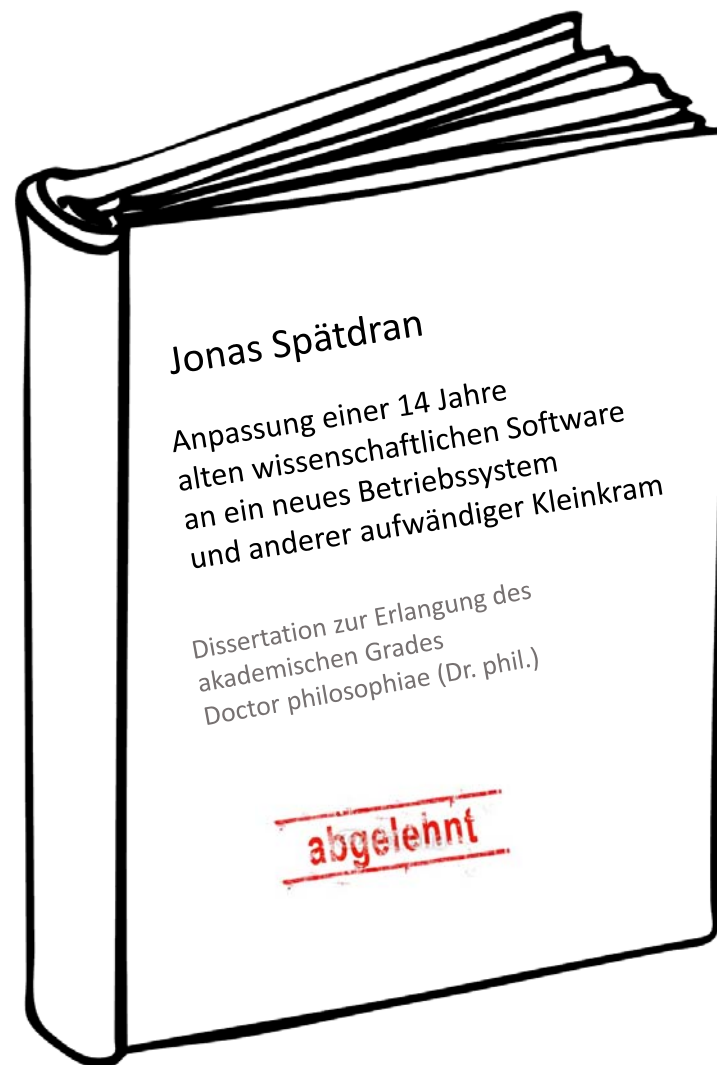
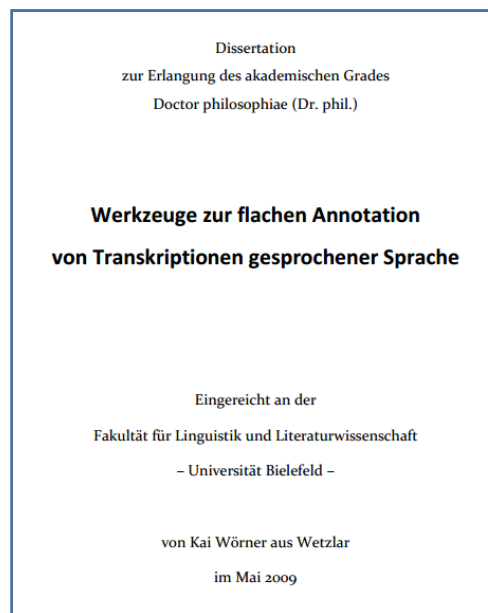
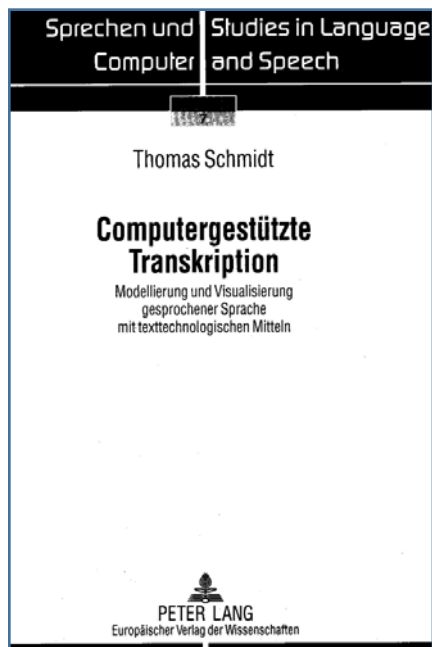


Neuer Drittmittelantrag?

- Keine dedizierten Förderprogramme für Entwicklung wissenschaftlicher Software
 - Softwareentwicklung muss im Antrag hinter anderer Fragestellung versteckt werden
- Maintenance ist keine Projektaufgabe
- Support ist keine Projektaufgabe
 - Dauerhafter Bedarf, nicht gut planbar
- Software-Entwicklung (in diesem Stadium) keine wissenschaftliche Tätigkeit (aber auch keine HiWi-Tätigkeit)
 - Entweder innovative Entwicklung oder Integration in laufendes System
 - Kaum geeignete KandidatInnen für Projektstellen



Bundesministerium
für Bildung
und Forschung



Kostenpflichtige Lizenzierung?

vgl. Qualitative Data Analysis

- MaxQDA – standard nicht-kommerzielle Lizenz, ca. 650 EURO
- atlas.ti, f4analyse, nvivo,
- Transana, Einzellizenz \$350

vgl. einfache („Consumer“) - Transkriptionsprogramme

- z.B. InqScribe - Voll-Lizenz \$99

Beispiele aus der Linguistik?

- Word Sketch Engine - £58 pro Jahr
- monoConc - \$85/2 Jahre
- wordSmith - £50

Marktanalyse, Business Plan, Lizenzmodell, Ausgründung

- ➔ Hohe Anfangsinvestition
- ➔ Rechtlich zulässig (jahrelange öffentliche Förderung)?
- ➔ Akzeptanz?
- ➔ Kunden außerhalb des akademischen „Marktes“?

„Kommerziell werden“

- ➔ Interoperabilität kann geschäftsschädigend sein
- ➔ Open Access-Idee in digitalen Infrastrukturen?
- ➔ Förderfähigkeit für Innovationen?

Das ideale Betriebsmodell...

- ✓ ... erlaubt weiterhin eine **kostenlose Distribution** der Software
- ✓ ... erlaubt weiterhin Entscheidungen nach Kriterien der beteiligten **wissenschaftlichen Communities**
- ✓ ... lässt einer **Leitung** ausreichend Raum für kurz-, mittel- und langfristige **Planungen**
- ✓ ... generiert finanzielle Mittel **bedarfsgerecht**
- ✓ ... generiert anteilig finanzielle Mittel für eine „**Grundfinanzierung**“
- ✓ ... erlaubt es, **Software-Entwickler** als solche zu bezahlen
- ✓ ... erlaubt es, **Support-Mitarbeiter** als solche zu bezahlen

Elemente

- ✓ Kostenpflichtige **Schulungen** für Lehrende, Forscherteams oder im Rahmen universitärer Lehrveranstaltungen
- ✓ Kostenpflichtige Unterstützung bei der **Aufbereitung (Kuration)** existierender Datenbestände für die Arbeit mit EXMARaLDA
- **Support-Verträge**
 - für **Fachbereiche**, die EXMARaLDA regelmäßig in der Lehre einsetzen
 - für **Forschungsprojekte**, die ihre Datengrundlage mit EXMARaLDA bearbeiten
 - für einzelne **Studierende und Promovierende**
- ✓ Partnerschaften mit **Forschungsprojekten**, die für ihre Arbeit spezielle **Erweiterungen** benötigen
- Partnerschaften mit **Fachverbänden** u.Ä., deren Mitglieder ein Interesse an der Weiterentwicklung von EXMARaLDA haben
- **Micro-Payment** auf freiwilliger Basis für einzelne Support-Fälle;
- **Crowd-Funding** für die Umsetzung aufwändiger Features, die für einen größeren Nutzerkreis von Interesse sind.
- **Eigene Drittmittelprojekte** für innovative Erweiterungen (bis Beta, danach Integration in übriges Betriebsmodell)

hzsk services

› hzsk services

› [Datenmanagement](#)

› [Schulungen](#)

› [Transkriptionsservice](#)

› [Mitarbeiter](#)

› [Referenzen](#)

› [Kontakt](#)

› [Impressum](#)

Linguistisches Datenmanagement für Forschung und Unternehmen

hask services bietet Beratung und technischen Support im Bereich des Datenmanagements, der Datenverarbeitung, -aufbereitung und -auswertung sowie der Erstellung und Verarbeitung sprachlicher Datenressourcen, insbesondere auch mündlicher Sprachkorpora, an. In individuellen Schulungen und Workshops vermitteln unsere Mitarbeiter methodologisches Fachwissen und technische Fertigkeiten für die erfolgreiche Durchführung von Forschungs- und Entwicklungsprojekten.

hask services ist ein an das [Hamburger Zentrum für Sprachkorpora \(HZSK\)](#) der Universität Hamburg angegliedertes, auf Wissens- und Technologietransfer spezialisiertes **Kompetenzzentrum** der [Hamburg Innovation GmbH](#).

Die Leistungen werden in der Regel als Gegenstand eines Werkvertrages, eines Forschungs- und Entwicklungsauftrages oder auf Honorarbasis erbracht.

2015 hzsk-services.de- This is a Free Drupal Theme

Ported to Drupal for the Open Source Community by [Drupalizing](#), a Project of [More than \(just\) Themes](#). Original design by [Simple Themes](#).

<http://hask-services.de/>

Preisschilder

EXMARaLDA
Download

0 EURO

1-tägige
EXMARaLDA
Schulung

500 EURO
+ RK

2-tägige
EXMARaLDA
Schulung

800 EURO
+ RK

projektspezifische
Erweiterung

50 EURO
Stundensatz

Support
Kontaktaufnahme

0 EURO

Support-Lizenz
Forschungsprojekt

? EURO

Support-Lizenz
Fachbereich

? EURO

Kuration 20h
komplexes
Videokorpus

15.000 EURO

Offene Fragen / Ungelöste Probleme

- Wie kommunizieren?
 - „Gestern noch 0 EURO, heute schon 800 EURO“
 - „Wir können nicht helfen, Sie haben nichts bezahlt“
- Anschubfinanzierung?
 - initiale Konzeption
 - Akquise
- Zielgröße?
- Rechtsform?

gwin

DFG

hzsk

Sonderforschungsbereich
Mehrsprachigkeit



INSTITUT FÜR
DEUTSCHE SPRACHE



CLARIN-D



Bundesministerium
für Bildung
und Forschung