

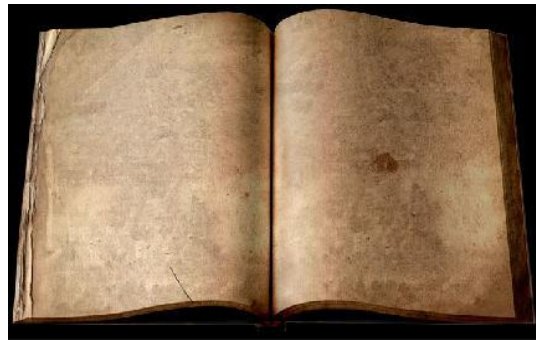
Wissens- und Content Management

Canonical Text Services (Allgemein)

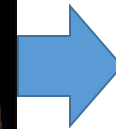
Dr. Jochen Tiepmar

Abteilung Automatische Sprachverarbeitung, Universität Leipzig

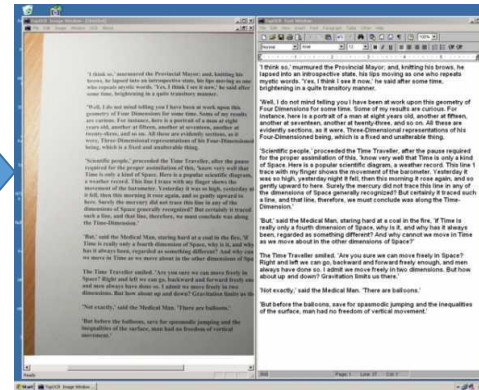
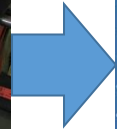
Vom Buch zur Datei



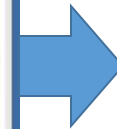
„Real“ book



Scans



OCR



Manuelles Editieren



Digital Edition

```

<tei:div n="2" type="strophe">
  <l n="1">O Christmas tree, O Christmas tree, </l>
  <l n="2">Much pleasure doth thou bring me ! </l>
  <l n="3">For every year the Christmas tree, </l>
  <l n="4">Brings to us all both joy and glee. </l>
  <l n="5">O Christmas tree, O Christmas tree, </l>
  <l n="6">Much pleasure doth thou bring me ! </l>
</tei:div>
<tei:div n="3" type="strophe">
  <l n="1">O Christmas tree, O Christmas tree, </l>
  <l n="2">Thy candles shine out brightly ! </l>
  <l n="3">Each bough doth hold its tiny light, </l>
  <l n="4">That makes each toy to sparkle bright. </l>
  <l n="5">O Christmas tree, O Christmas tree, </l>
  <l n="6">Thy candles shine out brightly ! </l>
</tei:div>

```

<http://free-download.ucoz.com/load/47-1-0-418>

https://en.wikipedia.org/wiki/File:Internet_Archive_book_scanner_1.jpg

<http://topocr.en.softonic.com/>

<http://www.jjc.edu/admissions/Pages/Returning-Students.aspx>

http://ctstest.informatik.uni-leipzig.de/demo/cts/?request=GetPassage&urn=urn:cts:songs:christmas.ohtennenbaum.en:&configuration=epidoc=true_divs=false

Datenquellen

Lokale Dateien

Dokumentenarchive (Online/Offline)


Quellcode Repositories (Github,...)

Forums, Social Media


Alles, was Text enthält

Auswahl an Online Archiven

RESEARCH CENTRE LANGUAGE
DEUTSCHES TEXTARCHIV (GERMAN TEXT ARCHIVE)



The German Text Archive (Deutsches Textarchiv, DTA) presents online a selection of key German-language works in various disciplines from the 17th to 19th centuries. The electronic full-texts are indexed linguistically and the search facilities tolerate a range of spelling variants.



The DTA presents German-language printed works from around 1650 to 1900 as full text and as digital facsimile. The selection of texts was made on the basis of lexicographical criteria and includes scientific or scholarly texts, texts from everyday life, and literary works. The digitalisation was made from the first edition of each work. Using the digital images of these editions, the text was first typed up manually twice ('double keying'). To represent the structure of the text, the electronic full-text was

momentan verboten | <SCRIPT>: 5 | <OBJECT>: 0

Finstellun

→ CONTACT

DR.
ALEXANDER GEYKEN
Arbeitsstellenleiter Digitales Wörterbuch der
deutschen Sprache
Tel: +49 (0)30 20370 390
→ geyken@bbaw.de
Jägerstraße 22-23
10117 Berlin

→ SYNOPSIS

→ STAFF

→ RESEARCH PROJECT

<http://www.bbaw.de/en/research/dta>

- German Text Archive from BBAW
- Textfiles
- Key german works from 17th to 19th century
- Scientific/scholarly works, everyday life and literature
- Tokenization, Lemmatization, Structurization and POS-Tagging
- 1300 editions planned, >1000 already included
- Text structure standardised

Auswahl an Online Archiven

The screenshot shows the Spiegel Online Kultur website. At the top, there is a navigation bar with links for Home, Video, Themen, Forum, English, DER SPIEGEL, SPIEGEL TV, Abo, and Shop. Below this is a search bar and a navigation menu with categories like Politik, Wirtschaft, Panorama, Sport, Kultur, Netzwelt, Wissenschaft, Gesundheit, einestages, Karriere, Uni, Reise, Auto, and Stil. A banner for 'Projekt Gutenberg-DE' is visible, featuring a portrait of Johannes Gutenberg. Below the banner is a search bar and a navigation menu with links for Startseite, Genres, Neue Texte, Alle Autoren, Alle Werke, Lesetips, Shop, Information, and Impressum. The main content area displays the title 'William Shakespeare: Hamlet, Prinz von Dänemark. (Übersetzer Wieland) - Kapitel 2' and the beginning of the text, including the characters Bernardo and Francisco.

Home | Video | Themen | Forum | English | DER SPIEGEL | SPIEGEL TV | Abo | Shop | Schlagzeilen | Wetter | TV-Programm | mehr ▼

SPIEGEL ONLINE KULTUR Login | Registrierung

Politik | Wirtschaft | Panorama | Sport | Kultur | Netzwelt | Wissenschaft | Gesundheit | einestages | Karriere | Uni | Reise | Auto | Stil

Ein Angebot von **PROJEKT GUTENBERG-DE**

Projekt Gutenberg-DE Textsuche bei Gutenberg-DE: Suchen

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 > William Shakespeare > Hamlet, Prinz von Dänemark. (Übersetzer Wieland)

William Shakespeare: Hamlet, Prinz von Dänemark. (Übersetzer Wieland) - Kapitel 2 [Quellenangabe](#)

Navigation: Kapitel 2

**Erster Aufzug.
Erste Scene.**

(Eine Terrasse vor dem Palast.)

(Bernardo und Francisco, zween Schildwachen, treten auf.)

Bernardo. Wer da?

Francisco. Nein, gebt Antwort: Halt, und sagt wer ihr seyd.

Bernardo. Lang lebe der König!

Francisco. Seyd ihr Bernardo?

Bernardo. Er selbst.

Francisco. Ihr kommt recht pünktlich auf eure Stunde.

Bernardo. Es hat eben zwölf geschlagen; geh du zu Bette, Francisco.

<http://gutenberg.spiegel.de/>

- Wachsende Sammlung deutscher Literatur ohne Copyright
- Textdateien
- >5000 Dokumente von >1100 Autoren
- Gemeinsames Korrekturlesen bei www.gaga.net
- Textstruktur heterogen

Auswahl an Online Archiven

HATHI TRUST
Digital Library

Search HathiTrust's digital library FULL-TEXT CATALOG

Search words about or within the items All Fields Search

[Advanced full-text search](#) | [Search tips](#) Full view only

[? Should I search catalog or full-text?](#)

[Not with a partner institution? »](#)

HathiTrust is a [partnership](#) of academic & research institutions, offering a collection of millions of titles digitized from libraries around the world.

WHAT CAN YOU DO WITH HATHITRUST?

- BROWSE COLLECTIONS**
Explore user-created [featured collections](#).
- READ BOOKS ONLINE**
Read millions of titles online — [like this one!](#)
- READ BOOKS ON THE GO**
Take the library's books anywhere with our [mobile website](#).
- DOWNLOAD BOOKS* & CREATE COLLECTIONS**
*requires institutional login

<https://www.hathitrust.org/>

- >3mio HD-Scans von Editionen aus verschiedenen Quellen
- Bilder oder PDF, kein Text (?)
- 30 Sprachen
- Login nötig für vollen Zugriff, ausgewählte Dokumente in Public Domain
- Werke aus Forschung, Lehre, Literatur, Politik,...

Auswahl an Online Archiven

The screenshot shows the homepage of the Wortschatz Portal. At the top, it says "Willkommen beim Wortschatz-Portal." Below this is a search bar with the label "Wort:" and a "Suche!" button. There is also a checkbox for "Beachte Groß-/Kleinschreibung". The logo for "WORTSCHATZ UNIVERSITÄT LEIPZIG" is prominently displayed. A paragraph of text explains that data is automatically collected from public sources and that the portal does not represent the views of the project. Below the main content, there are three featured sections: "Wörter des Tages" (Words of the Day), "Neues Wortschatzportal" (New Wortschatz Portal), and "Wortschatzportal (alt)" (Old Wortschatz Portal). At the bottom, there are three more sections: "Wörterbuch" (Dictionary), "Webservices" (Web Services), and "Feedback".

Willkommen beim Wortschatz-Portal.

Wort:

Suche! ?

Beachte Groß-/Kleinschreibung

Die Daten werden aus sorgfältig ausgewählten öffentlich zugänglichen Quellen automatisch erhoben. Die Beispielsätze werden automatisch ausgewählt und stellen keine Meinungsäußerung des Projektes Deutscher Wortschatz dar. Für die darin enthaltenen Inhalte und Meinungen sind ausschließlich die Autoren verantwortlich. Auch ohne besondere Kennzeichnung unterliegen im Wortschatz wiedergegebene Marken wie Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. den gesetzlichen Bestimmungen. Die synonyme Verwendung eines Trademarks beschreibt nicht notwendigerweise produktspezifische Eigenschaften sondern kennzeichnet stattdessen die Verwendung des Begriffs im allgemeinsprachlichen Kontext.

Wörter des Tages

Die tagesaktuellen Begriffe. Ausgewählt aus Tageszeitungen und Newsdiensten. Täglich um 7 Uhr früh. Seit April 2002. Jetzt auch als RSS 2.0!

Neues Wortschatzportal

Das neue internationale Wortschatzportal bietet erweiterte Zugriffsmöglichkeiten auf die Korpora der Leipzig Corpora Collection (LCC). Derzeit stehen mehr als 230 Korpora in über 200 Sprachen zur Verfügung. Das Portal wird kontinuierlich weiter entwickelt und um neue Ressourcen erweitert.

Wortschatzportal (alt)

Auf unserem internationalen Wortschatzportal in englischer Sprache können Sie derzeit in Korpora 229 verschiedener Sprachen Wörter nachschlagen. *Dieses Portal wird nicht mehr weiter entwickelt.*

Wörterbuch

Über 100.000 Wörter und Wendungen auf Deutsch und Englisch. Die

Webservices

Mit den Webservices ist ein direkter Zugriff auf die Daten des Projektes

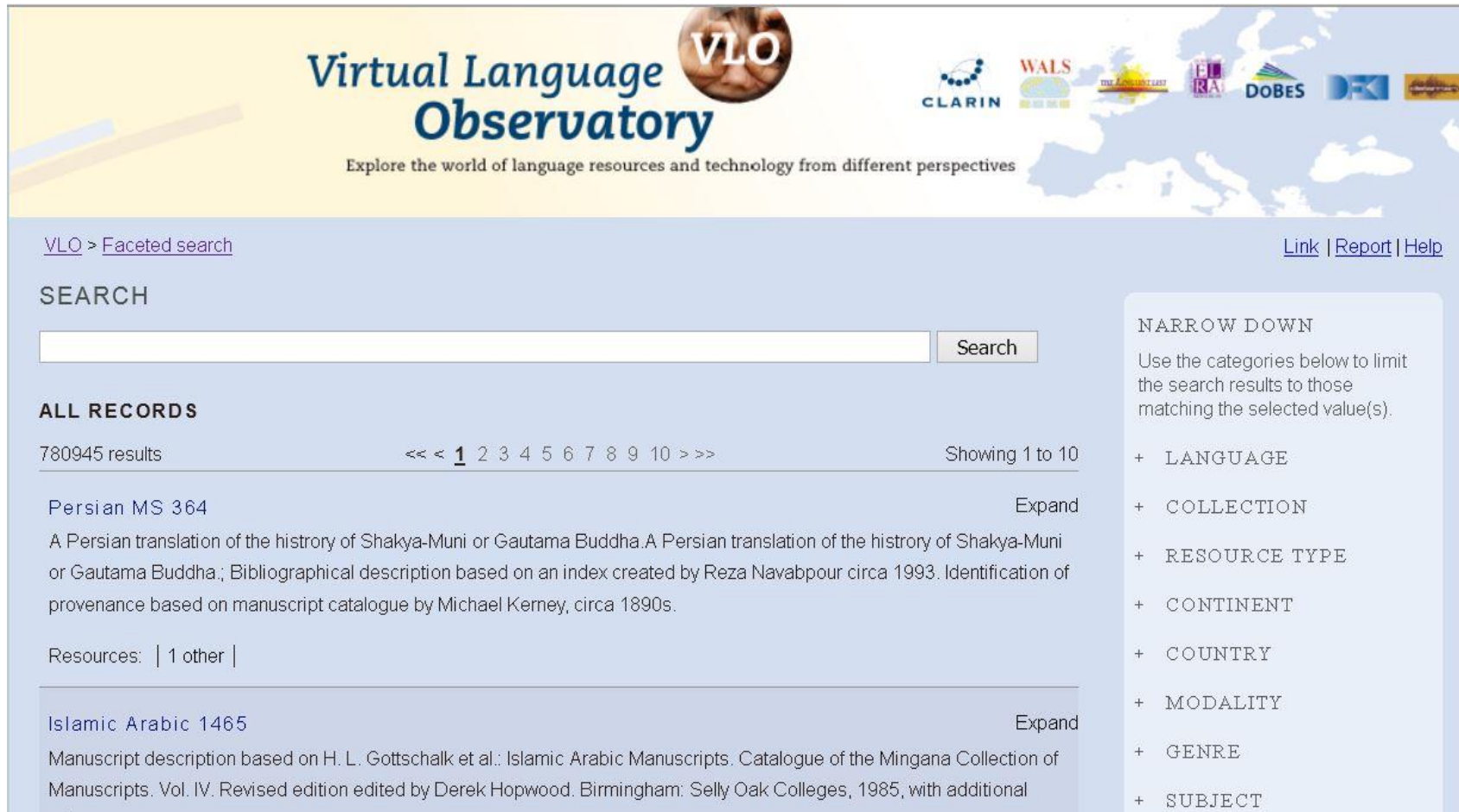
Feedback

Feedback der Nutzer zu den verschiedenen Services des

<http://wortschatz.uni-leipzig.de/>

- Wort als Ressource
- Wortstatistiken, Trend Analysen, Beispielsätze,...
- Laufend aktualisiert auf Basis ausgewählter Online Ressourcen

Auswahl an Online Archiven



Virtual Language Observatory
Explore the world of language resources and technology from different perspectives

VLO > [Faceted search](#) [Link](#) | [Report](#) | [Help](#)

SEARCH

ALL RECORDS

780945 results << < **1** 2 3 4 5 6 7 8 9 10 > >> Showing 1 to 10

Persian MS 364 [Expand](#)
A Persian translation of the history of Shakya-Muni or Gautama Buddha. A Persian translation of the history of Shakya-Muni or Gautama Buddha.; Bibliographical description based on an index created by Reza Navabpour circa 1993. Identification of provenance based on manuscript catalogue by Michael Kerney, circa 1890s.
Resources: | 1 other |

Islamic Arabic 1465 [Expand](#)
Manuscript description based on H. L. Gottschalk et al.: Islamic Arabic Manuscripts. Catalogue of the Mingana Collection of Manuscripts. Vol. IV. Revised edition edited by Derek Hopwood. Birmingham: Selly Oak Colleges, 1985, with additional

NARROW DOWN
Use the categories below to limit the search results to those matching the selected value(s).

- + LANGUAGE
- + COLLECTION
- + RESOURCE TYPE
- + CONTINENT
- + COUNTRY
- + MODALITY
- + GENRE
- + SUBJECT

<https://vlo.clarin.eu/?7>

Facettierte
Dokumentensuche
- Suchraum kann facettiert
eingeschränkt werden
(genre, modality,
organisation,...)
- 780945 Dokumente

Auswahl an Online Archiven

<http://www.perseus.tufts.edu/hopper/>

<http://paralleltext.info/data/>

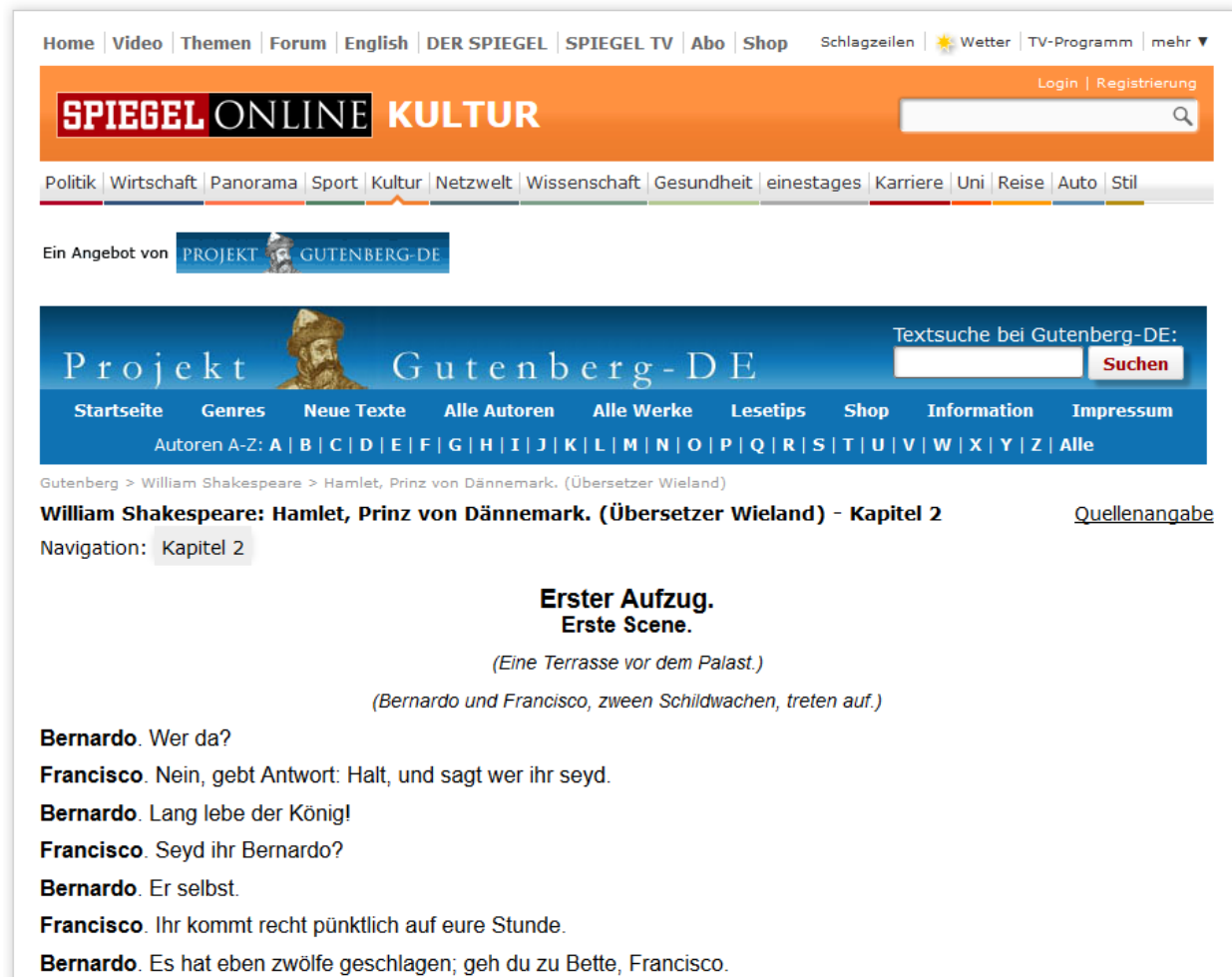
<http://shakespeare.mit.edu/>

<http://historicaltextarchive.com/>

<http://classics.mit.edu/>

(...)

Text als Webresource



Home | Video | Themen | Forum | English | DER SPIEGEL | SPIEGEL TV | Abo | Shop | Schlagzeilen | Wetter | TV-Programm | mehr ▼

SPIEGEL ONLINE KULTUR Login | Registrierung

Politik | Wirtschaft | Panorama | Sport | Kultur | Netzwelt | Wissenschaft | Gesundheit | einestages | Karriere | Uni | Reise | Auto | Stil

Ein Angebot von **PROJEKT GUTENBERG-DE**

Projekt Gutenberg-DE Textsuche bei Gutenberg-DE: **Suchen**

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 > William Shakespeare > Hamlet, Prinz von Dänemark. (Übersetzer Wieland)

William Shakespeare: Hamlet, Prinz von Dänemark. (Übersetzer Wieland) - Kapitel 2 [Quellenangabe](#)

Navigation: [Kapitel 2](#)

**Erster Aufzug.
Erste Scene.**

(Eine Terrasse vor dem Palast.)

(Bernardo und Francisco, zween Schildwachen, treten auf.)

Bernardo. Wer da?

Francisco. Nein, gebt Antwort: Halt, und sagt wer ihr seyd.

Bernardo. Lang lebe der König!

Francisco. Seyd ihr Bernardo?

Bernardo. Er selbst.

Francisco. Ihr kommt recht pünktlich auf eure Stunde.

Bernardo. Es hat eben zwölfe geschlagen; geh du zu Bette, Francisco.

<http://gutenberg.spiegel.de/buch/hamlet-prinz-von-dannemark-ubersetzer-wieland-5597/2>

Projektspezifische Webseiten

Keine offensichtliche Möglichkeit zum automatisierten Abfragen

Crawlen möglicherweise kompliziert
(Requestlimits, Werbung, Javascript,...)

Hamlet von von Schlegel(Project Gutenberg)

```
<div class="chapter" id="chap01">
  <h3>Erster Aufzug.</h3>
  <h4>Erste Scene.</h4>
  <p class="center">
    <span class="regie">(Eine Terrasse vor dem Palast.)</span>
  </p>
  <p class="center">
    <span class="regie">(Bernardo und Francisco, zween Schildwachen, treten auf.)</span>
  </p>
  <p>
    <span class="speaker">Bernardo</span>. Wer da?</p>
  <p>
    <span class="speaker">Francisco</span>. Nein, gebt Antwort: Halt, und sagt wer ihr seyd.</p>
  <p>
    <span class="speaker">Bernardo</span>. Lang lebe der König!</p>
  <p>
```

Hamlet von Wieland (Project Gutenberg)

```
<div class="chapter" id="chap01">  
  <h3>ERSTER AKT</h3>  
  <p class="center">ERSTE SZENE  
  <br />  
  <br />  
  <i>Helsingör. Eine Terrasse vor dem Schlosse</i></p>  
  <p>  
  <br />  
  <i>Francisco auf dem Posten, Bernardo tritt auf.</i></p>  
  <p>BERNARDO  
  <br /> Wer da?</p>  
  <p>FRANCISCO  
  <br /> Nein,  
  <em>mir</em> antwortet; steht und gebt Euch kund!</p>  
  <p>BERNARDO  
  <br /> Lang lebe der König!</p>
```

Hamlet von Moby

<h3>SCENE I. Elsinore. A platform before the castle.</H3>

<p><blockquote>

<i>FRANCISCO at his post. Enter to him BERNARDO</i>

</blockquote>

BERNARDO

<blockquote>

Who's there?

</blockquote>

FRANCISCO

<blockquote>

Nay, answer me: stand, and unfold yourself.

</blockquote>

BERNARDO

<blockquote>

Long live the king!

</blockquote>

Hamlet von Folger

```
<stage xml:id="stg-0000" n="SD 1.1.0" type="entrance" who="#Barnardo_Ham #Francisco_Ham">
<w xml:id="w0000070" n="SD 1.1.0">Enter</w>
<c xml:id="c0000080" n="SD 1.1.0"> </c>
<w xml:id="w0000090" n="SD 1.1.0">Barnardo</w>
<c xml:id="c0000100" n="SD 1.1.0"> </c>
<w xml:id="w0000110" n="SD 1.1.0">and</w>
<c xml:id="c0000120" n="SD 1.1.0"> </c>
<w xml:id="w0000130" n="SD 1.1.0">Francisco</w>
<pc xml:id="p0000140" n="SD 1.1.0">,</pc>
<c xml:id="c0000150" n="SD 1.1.0"> </c>
<w xml:id="w0000160" n="SD 1.1.0">two</w>
<c xml:id="c0000170" n="SD 1.1.0"> </c>
<w xml:id="w0000180" n="SD 1.1.0">sentinels</w>
<pc xml:id="p0000190" n="SD 1.1.0">.</pc>
</stage>
<lb xml:id="lb-00003"/>
<lb xml:id="lb-00004"/>
<sp xml:id="sp-0001" who="#Barnardo_Ham">
<speaker xml:id="spk-0001">
<w xml:id="w0000200">BARNARDO</w>
</speaker>
<ab xml:id="ab-0001">
<milestone unit="ftln" xml:id="ftln-0001" n="1.1.1" ana="#prose" corresp="#w0000210 #c0000220 #w0000230 #p0000240"/>
<w xml:id="w0000210" n="1.1.1">Who's</w>
<c xml:id="c0000220" n="1.1.1"> </c>
<w xml:id="w0000230" n="1.1.1">there</w>
<pc xml:id="p0000240" n="1.1.1">?</pc>
```

Daten Heterogenität

Kein gemeinsames Text Format & Zugriff

Keine Abgrenzung von Textstruktur, Metainformation & Text

<speech> vs <speaker> vs Elliot:

Dateien und Workflows über Projektgrenzen hinweg nicht kompatibel

TEI/XML

<https://github.com/PerseusDL/canonical-greekLit/blob/master/data/tlg0020/tlg003/tlg0020.tlg003.perseus-grc1.xml>

```
<teiheader> ... </teiheader>
<text id="stoa0045.stoa0"><body>
<div1 n="Sh." type="Book">
<milestone ed="P" n="1" unit="card"/>
<l>
<milestone ed="P" unit="para"/>ἡ οἴη προλιποῦσα δόμους καὶ πατρίδα γαῖαν</l>
<l>ἦλυθεν ἐς Θήβας μετ' ἀρήιον Ἀμφιτρύωνα</l>
<l>Ἀλκμήνη, θυγάτηρ λαοσσόου Ἡλεκτρύωνος</l>
<l>ἡ ῥα γυναικῶν φύλον ἐκαίνυτο θηλυτεράων</l>
<l n="5">εἶδει τε μεγέθει τε· νόον γε μὲν οὐ τις ἔριζε</l>
<l>τάων, ἄς θνηταὶ θνητοῖς τέκον εὐνηθεῖσαι.</l>
```

TEI/XML

<https://github.com/PerseusDL/canonical-greekLit/blob/master/data/tlg0020/tlg003/tlg0020.tlg003.perseus-eng1.xml>

```
<teiheader> ... </teiheader>
```

```
<text><body>
```

```
<div1 n="Sh." type="book">
```

```
<div2 type="card" n="1"><p>
```

Or like her who left home and country and came to Thebes, following warlike Amphitryon, — even Alcmena, the daughter of Electryon, gatherer of the womankind <milestone unit="line" n="5" ed="p"/> in beauty and in height; and in wisdom none vied with her of those whom mortal women bore of union with such charm as comes from golden Aphrodite. And she so honored her husband in her heart <milestone unit="line" n="10" ed="p"/> as none of womankind did violently when he was angry about oxen; so he left his own country and came to Thebes and was suppliant to the shield-carrying men of Cadmus. There he <milestone unit="line" n="15" ed="p"/> without the joys of love, nor might he go in unto the neat-ankled daughter of Electryon until he had avenged the death of his wife's great-husband, the villages of the heroes, the Taphians and Teleboans; <milestone unit="line" n="20" ed="p"/> for this thing was laid upon him, and the gods were witnesses to perform the great task to which Zeus had bound him. With him went the horse-driving Boeotians, breathing above their shields, <milestone unit="line" n="25" ed="p"/> on their hand, and the gallant Phocians eager for war and battle. And the noble son of Alcaeus led them, rejoicing in his host.

But the father of men and gods was forming another scheme in his heart, to beget one to defend against destruction gods and men who eat bread. <milestone unit="line" n="30" ed="p"/> <placeName key="tgn,7011019">Olympus</placeName> by night pondering guile in the deep of his heart, and yearned for the love of the well-girded woman. Again wise Zeus went on and trod the highest peak of Phicium: <note resp="Loeb" anchored="yes">A mountain peak near <placeName key="perseus,Thebes">Thebes</placeName> the Sphinx & lpar; called in <bibl n="Hes. Th. 326">Hes. Th. 326</bibl> <foreign lang="greek">fi=c</foreign> & rpar; .</note> there he sat and planned marvels. <milestone unit="line" n="35" ed="p"/> So in one night Zeus shared the bed and love of the neat-ankled daughter of Electryon and fulfilled his desire; and in the same night Amphitryon, gathered his house when he had ended his great task. </p></div2>

Kanonische Zitation

- Referenzierung von Textstellen ist grundlegendes Werkzeug in akademischen Arbeiten
 - Üblicherweise anhand von Seitenzahlen
- Seitenzahlen bei digitalen Dokumenten unpraktisch
 - Text unabhängig vom physischen Medium
- Inhaltliche Strukturelemente als Ankerpunkte
 - Kapitel, Buch, Strophe, ...
- -> Strukturbasiertes Referenzierungssystem benötigt
- Zitatspuren als akademisches Zitat oder Forschungsgegenstand

Canonical Text Services

- Webserviceprotokoll zur Referenzierung von Textstellen auf Basis der Dokumentenstruktur (unabhängig vom physischen Medium)
- Entwickelt im Homer Multitext Projekt
 - Neel Smith & Christopher Blackwell
- Referenzen über Unique Resource Names (URNs)
- Technologieunabhängig, systemübergreifend
- Ähnlich ISBN oder DOI, aber flexible Granularität und Dynamik
 - Komputational komplex

like “Rinder” and “Esel”, which occur often. In elberfelder1905, these numbers are spelled in full⁷.

Because of the connection to CTS, each of the citations is referenced as a CTS URN and can be used to retrieve the corresponding text passage. A simple visualization based on a prototype of (Reckziegel et al., 2016)’s CTRaCE uses this connection to list each citation with links to the corresponding text passage in DTA:

```
Am Anfang schuf Gott Himmel und Erde .  
urn:cts:pb:deu.luther1545:1.1.1  
-- am anfang schuf gott himmel und erde  
-- im anfang schuf gott himmel und erde8
```

Furthermore, since PBC is a parallel corpus, the URNs can be used to align the citations over different translations in one language, for example resulting from different use of metaphors. The text passage “ich wache, und bin wie ein einsamer Vogel auf dem Dache.” from elberfelder1871 was found as a citation in DTA. The corresponding text passage “Ich bin gleich wie eine Rohrdrommel in der Wüste; ich bin gleich wie ein Käuzlein in den verstörten Stätten.” in luther1545 was not found but can be associated when the results for the different translations are aligned using the alignment of CTS URNs. Furthermore, this means that any

⁷ See [urn:cts:pb:deu.elberfelder1905:4.31.34](#) and [urn:cts:pb:deu.elberfelder1905:4.31.33](#)

Begriffe

URI

Unique Resource Identifier
Name einer Resource oder Ort im Internet

URL

Unique Resource Location
Ort im Internet

URN

Unique Resource Name
Name einer Resource im Internet

(<https://www.w3.org/TR/uri-clarification/>)

Begriffe

CTS URN

Unique Resource Name für Canonical Text Service

Name eines spezifischen Textabschnitts in einem Canonical Text Service

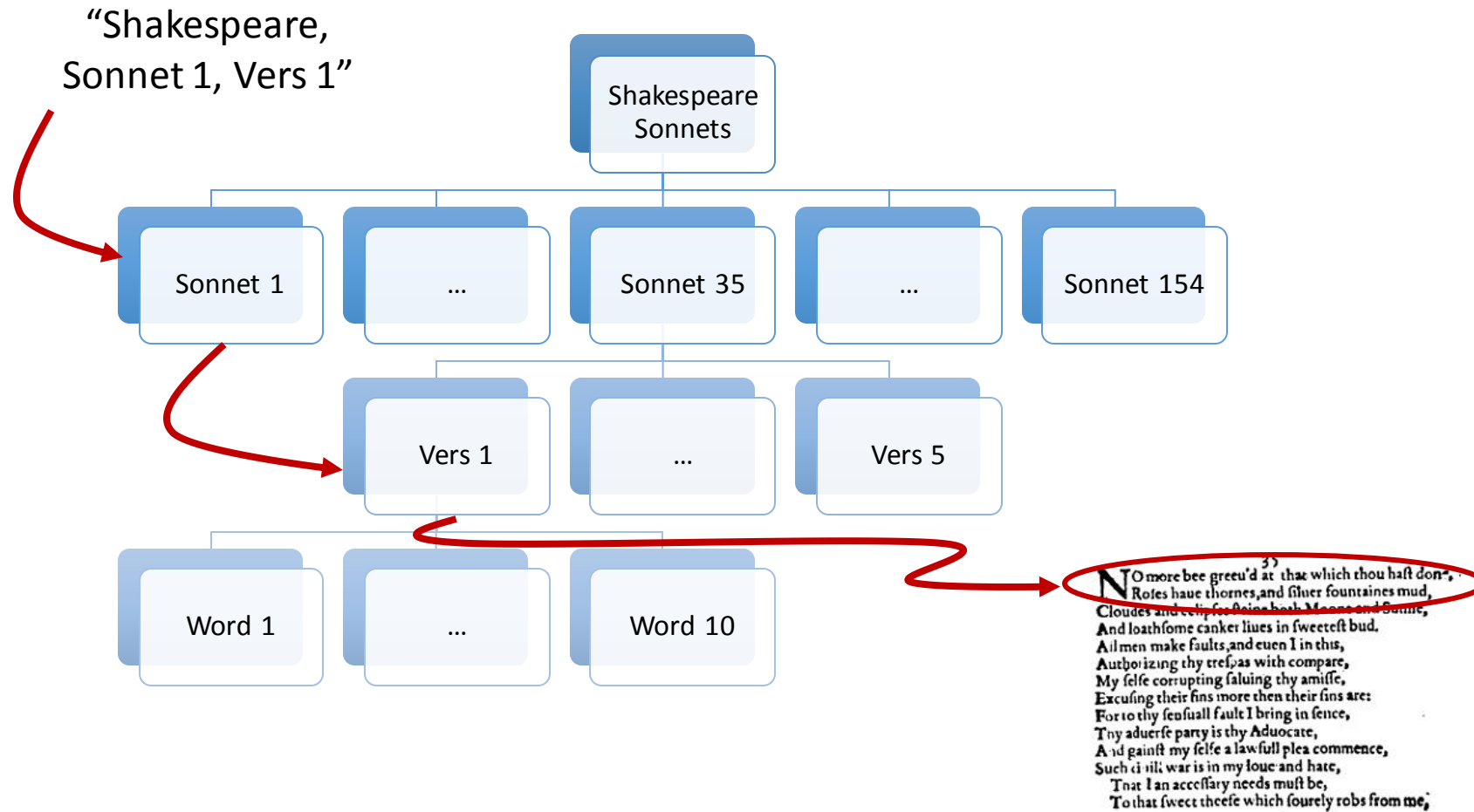
Text Inventory

Übersicht über CTS URNs und Metainformation auf Dokumentenebene

Textabschnitt

1 zusammenhängender Teil eines Textes

Documents Hierarchie



Zitation

Document „Äußere Hierarchie“

Shakespeare → Sonnets → english → 1st edition

Textabschnitt „Innere Hierarchie“

Sonnet 1 → Vers 1

Kombiniert

Shakespeare → Sonnets → english → 1st edition → Sonnet 1 → Vers 1

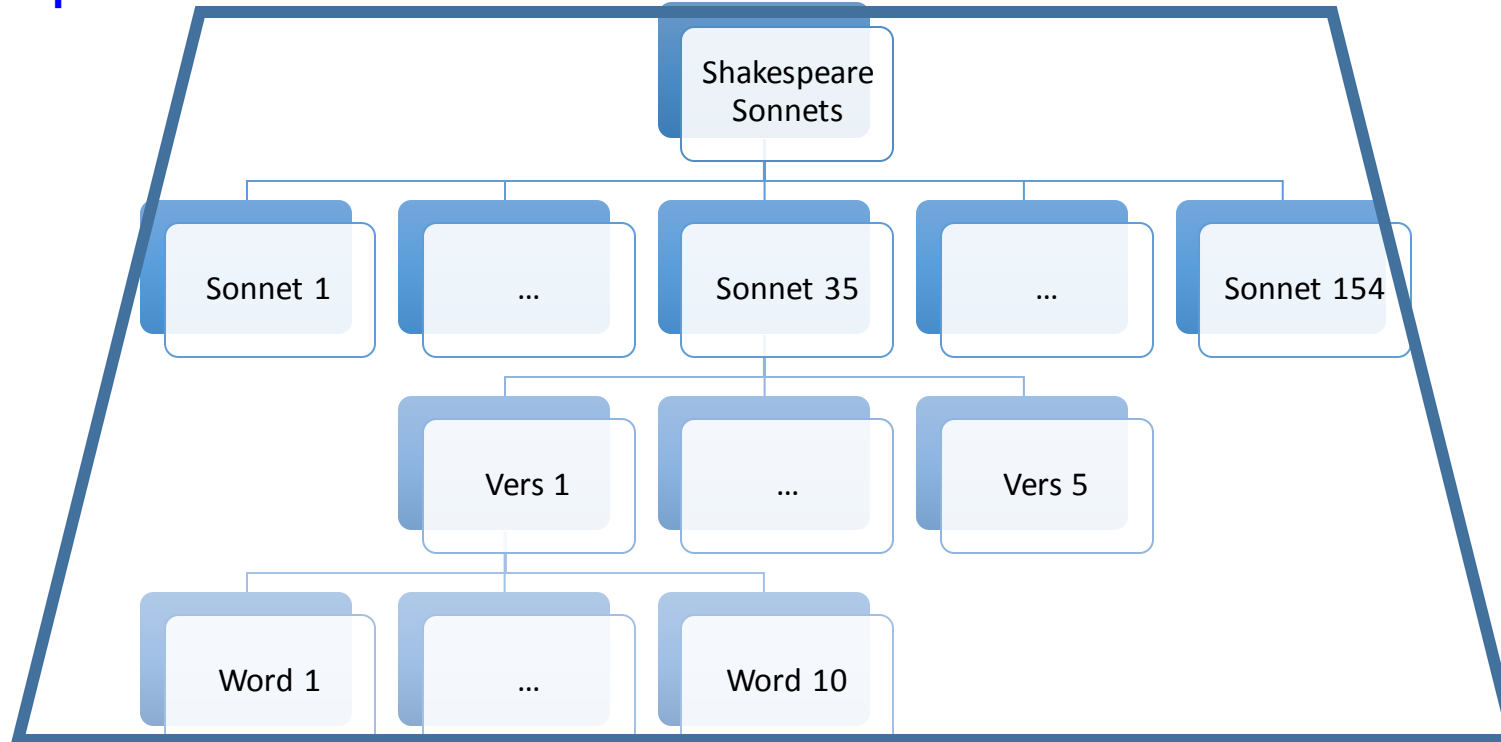
CTS-URN

urn:cts:demo:shakespeare.sonnets.en.1:1.1

Kanonische Zitation

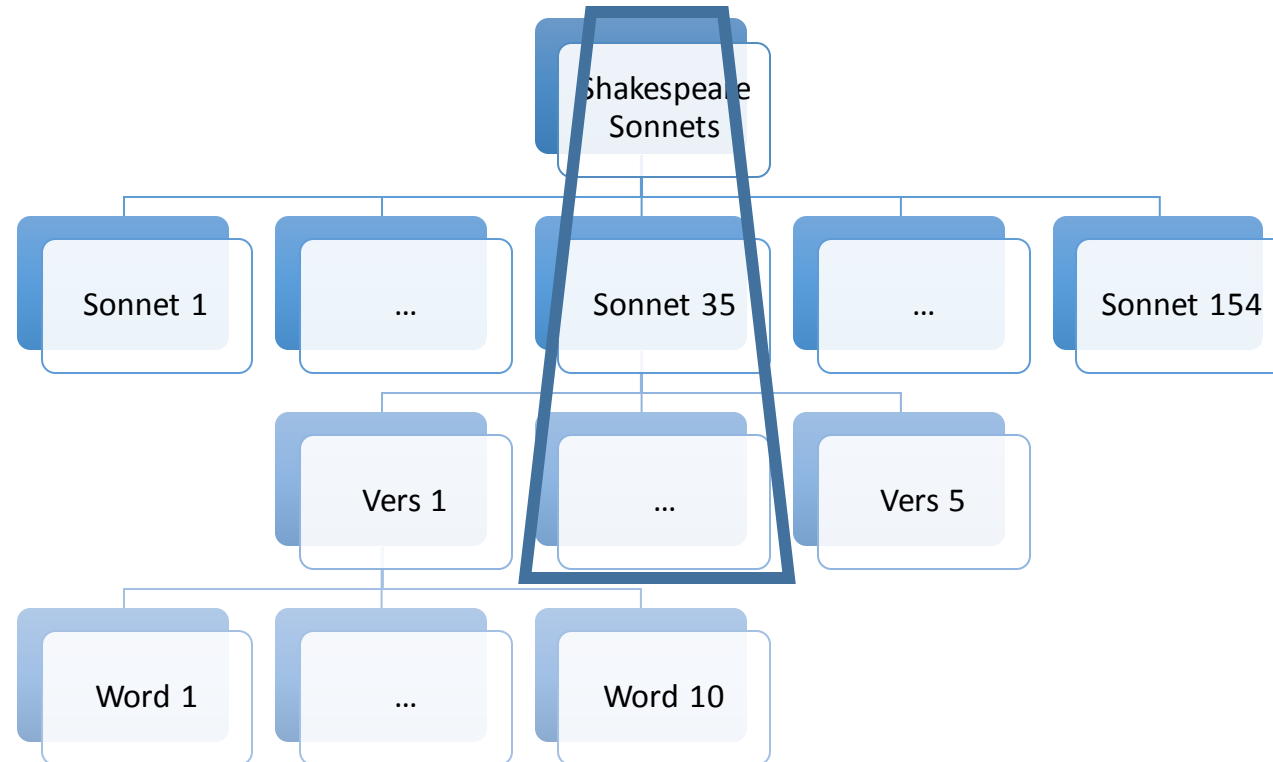
urn:cts:demo:shakespeare.sonnets:

urn:cts:demo:shakespeare.sonnets.de:



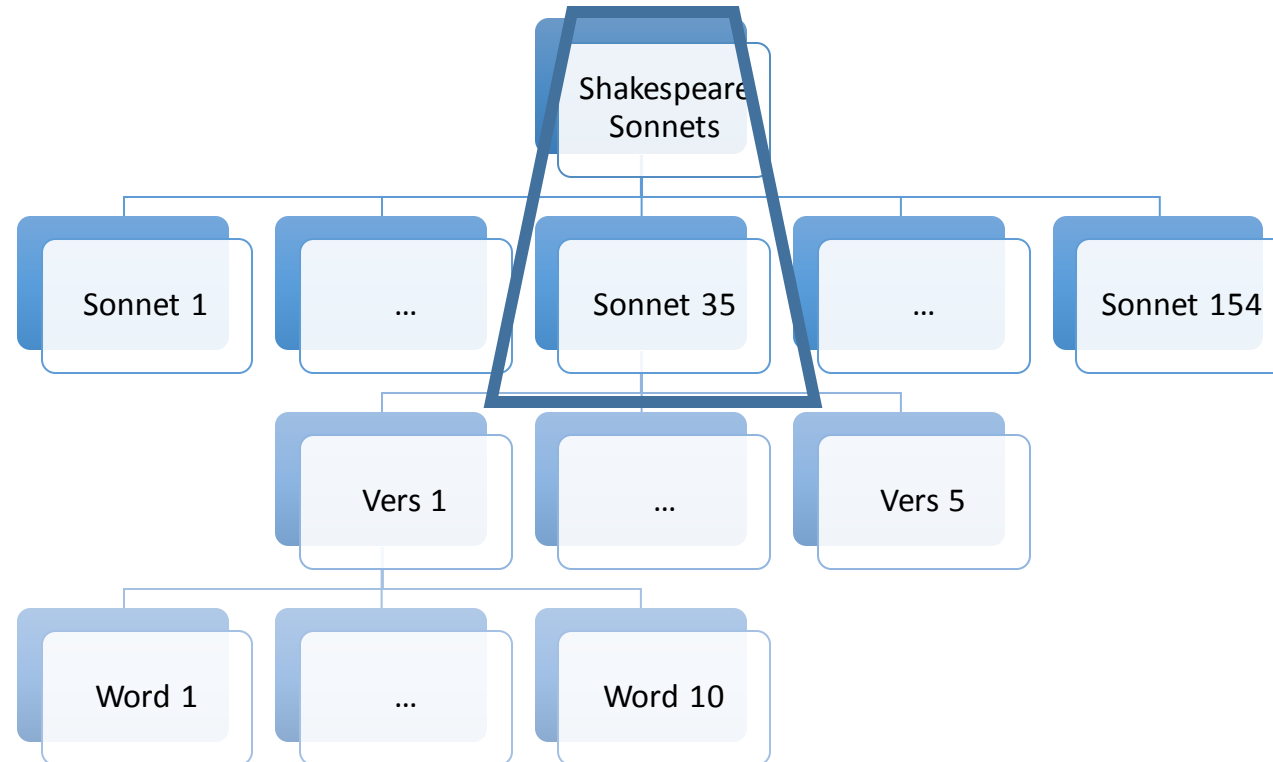
Kanonische Zitation

urn:cts:demo:shakespeare.sonnets:35.4



Kanonische Zitation

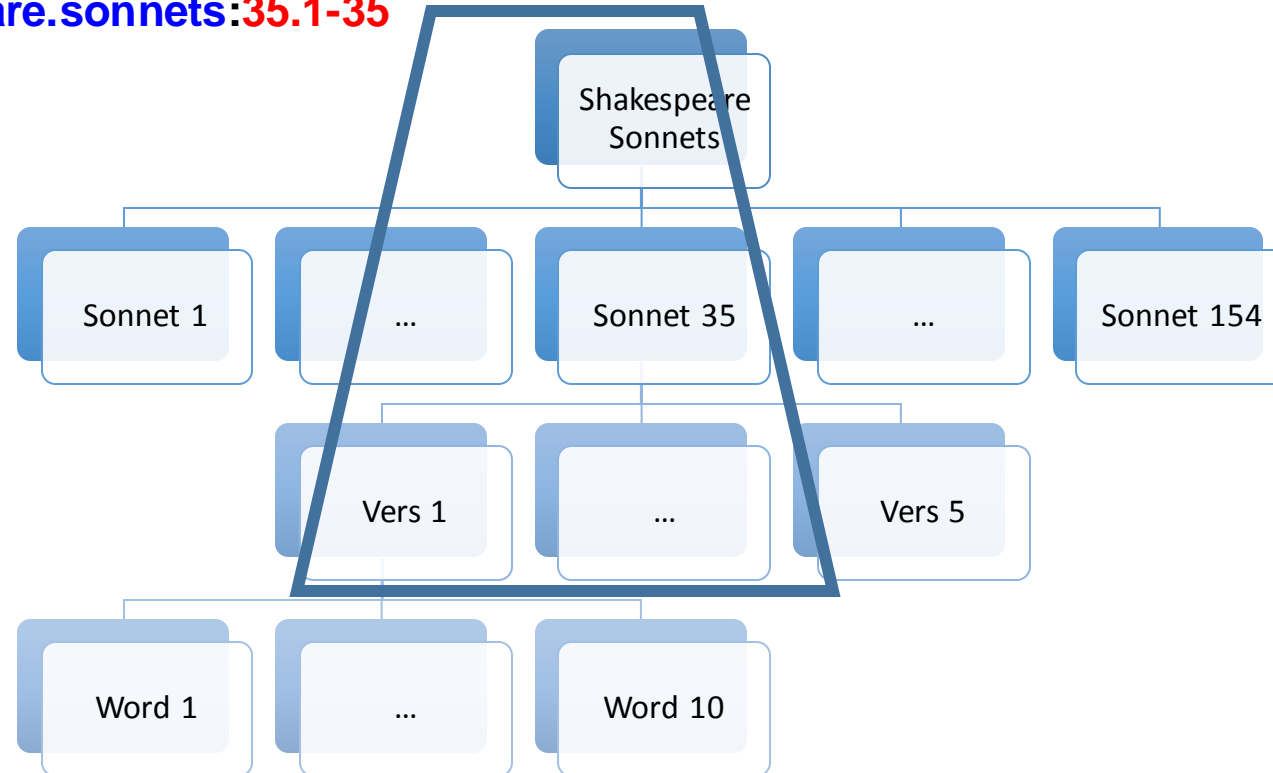
urn:cts:demo:shakespeare.sonnets:35



Kanonische Zitation

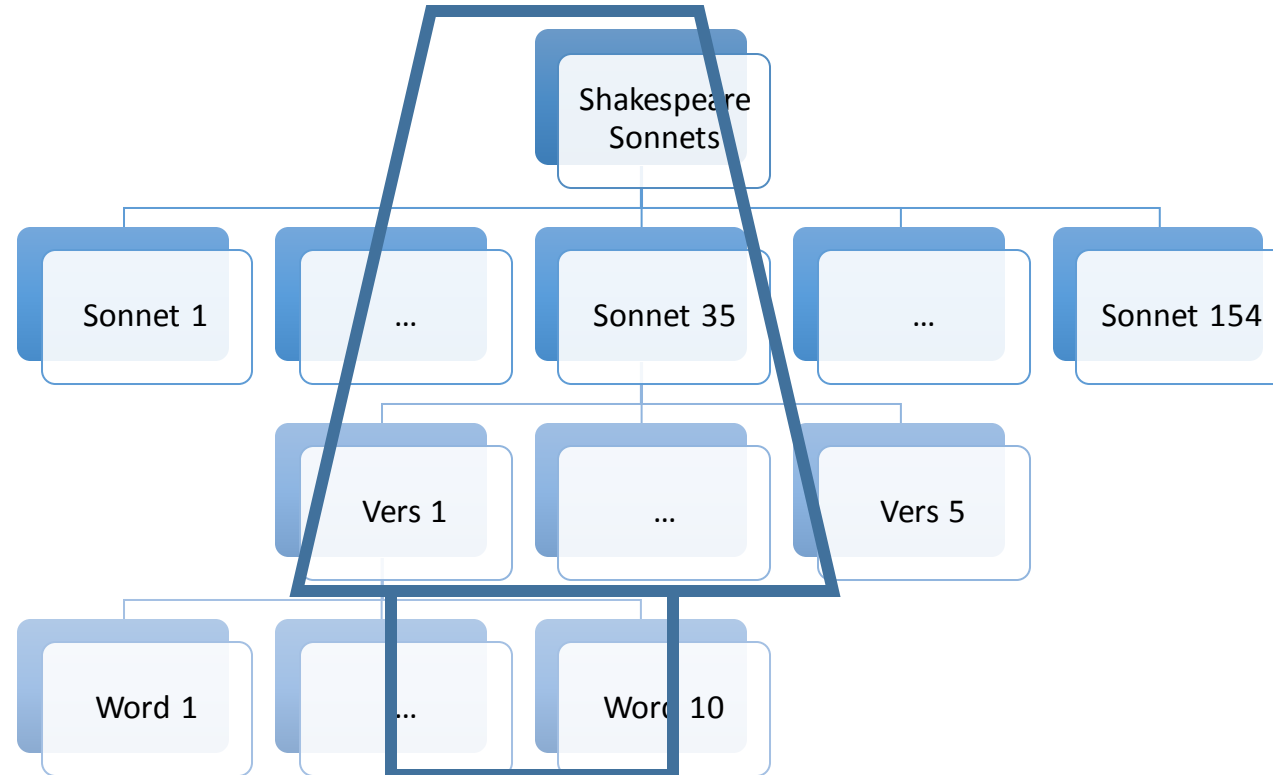
urn:cts:demo:shakespeare.sonnets:35.1-35.5

urn:cts:demo:shakespeare.sonnets:35.1-35



Kanonische Zitation

urn:cts:demo:shakespeare.sonnets:35.1@grieved-35.5@faults[1]



CTS URNs

urn:cts:demo:shakespeare.sonnets:35.1@grieved-35.5@faults[1]

urn:cts:[CTSNAMESPACE]:[WORK]:[PASSAGE]

[CTSNAMESPACE] = Namespace des Datensatzes

[WORK] = TEXTGROUP.WORK.VERSION.EXEMPLAR

[PASSAGE] = Textabschnitt

Zahlen sind Strings

-> Zahlen spiegeln nicht zwangsläufig Dokumentenreihenfolge wider

CTS URNs

urn:cts:demo:shakespeare.sonnets:35.1@grieved-35.5@faults[1]

35.5@vaults[1] == 35.5@vaults

urn:cts:, namespace and textgroup benötigt, alles weitere optional

Groß- und kleinschreibung entscheidend(latinlit vs. LatinLit)

Beschreibende Teile können alle erlaubten Zeichen enthalten, auch Zahlen

Korrekt:**urn:cts:demo:shakespeare.sonnets12.2:preface.line1**

chapter2.line4 kann vorkommen vor **chapter2.line1**

Funktionstragende Zeichen dürfen nicht beschreibend verwendet werden

. : - @ [] (complete list <http://www.homermultitext.org/hmt-docs/specifications/ctsur/>)

Inkorrekt:**urn:cts:demo:sha[@kespeare.sonn-ets:35.1[@-**

CTS URNs / Übung

- Sind diese URNs valide? Warum?
urn:cts:demo:goethe.faust.de:35.1

urn:cts:demo:shakespeare.sonnets:35.1@grieved-35.5@faults[1]

35.5@vaults[1] == 35.5@vaults

urn:cts:, namespace and textgroup benötigt, alles weitere optional

Groß- und kleinschreibung entscheidend(latinlit vs. LatinLit)

Beschreibend Teile können alle erlaubten Zeichen enthalten, auch Zahlen

Korrekt:urn:cts:demo:shakespeare.sonnets12.2:preface.line1

chapter2.line4 kann vorkommen vorchapter2.line1

Funktionstragende Zeichen dürfen nicht beschreibend verwendet werden

. :- @ []

(complete list <http://www.homermultitext.org/hmt-docs/specifications/ctsum/>)

Inkorrekt:urn:cts:demo:sha[@kespeare.sonn-ets:35.1]@-

CTS URNs / Übung

- Sind diese URNs valide? Warum?

urn:cts:demo:goethe.faust.de:35.1

Urn:cts:demo:goethe.faust.de:35.1

yes

urn:cts:demo:shakespeare.sonnets:35.1@grieved-35.5@faults[1]

35.5@vaults[1] == 35.5@vaults

urn:cts:, namespace and textgroup benötigt, alles weitere optional

Groß- und kleinschreibung entscheidend(latinlit vs. LatinLit)

Beschreibend Teile können alle erlaubten Zeichen enthalten, auch Zahlen

Korrekt:urn:cts:demo:shakespeare.sonnets12.2:preface.line1

chapter2.line4 kann vorkommen vorchapter2.line1

Funktionstragende Zeichen dürfen nicht beschreibend verwendet werden

. :- @ []

(complete list <http://www.homermultitext.org/hmt-docs/specifications/ctsum/>)

Inkorrekt:urn:cts:demo:sha[@kespeare.sonn-ets:35.1]@-

CTS URNs / Übung

- Sind diese URNs valide? Warum?

urn:cts:demo:goehte.faust.de:35.1

Urn:cts:demo:goehte.faust.de:35.1

urn:cts:Demo:goehte.faust.de:35.1

yes

urn:cts:demo:shakespeare.sonnets:35.1@grieved-35.5@faults[1]

35.5@vaults[1] == 35.5@vaults

no

urn:cts:, namespace and textgroup benötigt, alles weitere optional

Groß- und kleinschreibung entscheidend(latinlit vs. LatinLit)

Beschreibend Teile können alle erlaubten Zeichen enthalten, auch Zahlen

Korrekt:urn:cts:demo:shakespeare.sonnets12.2:preface.line1

chapter2.line4 kann vorkommen vorchapter2.line1

Funktionstragende Zeichen dürfen nicht beschreibend verwendet werden

. :- @ []

(complete list <http://www.homermultitext.org/hmt-docs/specifications/ctsum/>)

Inkorrekt:urn:cts:demo:sha[@kespeare.sonn-ets:35.1]@-

CTS URNs / Übung

- Sind diese URNs valide? Warum?

urn:cts:demo:goehte.faust.de:35.1

Urn:cts:demo:goehte.faust.de:35.1

urn:cts:Demo:goehte.faust.de:35.1

urn:cts:demo:goehte.faust.de:invalid35.1

yes

no

yes

urn:cts:demo:shakespeare.sonnets:35.1@grieved-35.5@faults[1]

35.5@vaults[1] == 35.5@vaults

urn:cts:, namespace and textgroup benötigt, alles weitere optional

Groß- und kleinschreibung entscheidend(latinlit vs. LatinLit)

Beschreibend Teile können alle erlaubten Zeichen enthalten, auch Zahlen

Korrekt:urn:cts:demo:shakespeare.sonnets12.2:preface.line1

chapter2.line4 kann vorkommen vorchapter2.line1

Funktionstragende Zeichen dürfen nicht beschreibend verwendet werden

. :- @ []

(complete list <http://www.homermultitext.org/hmt-docs/specifications/ctsum/>)

Inkorrekt:urn:cts:demo:sha[@kespeare.sonn-ets:35.1[@-

CTS URNs / Übung

- Sind diese URNs valide? Warum?

urn:cts:demo:goehte.faust.de:35.1

yes

Urn:cts:demo:goehte.faust.de:35.1

no

urn:cts:Demo:goehte.faust.de:35.1

yes

urn:cts:demo:goehte.faust.de:invalid35.1

yes

urn:cts:demo:goehte.faust.de:not-valid35.1

urn:cts:demo:shakespeare.sonnets:35.1@grieved-35.5@faults[1]

35.5@vaults[1] == 35.5@vaults

urn:cts:, namespace and textgroup benötigt, alles weitere optional

Groß- und kleinschreibung entscheidend(latinlit vs. LatinLit)

Beschreibend Teile können alle erlaubten Zeichen enthalten, auch Zahlen

Korrekt:urn:cts:demo:shakespeare.sonnets12.2:preface.line1

chapter2.line4 kann vorkommen vorchapter2.line1

Funktionstragende Zeichen dürfen nicht beschreibend verwendet werden

. :- @ []

(complete list <http://www.homermultitext.org/hmt-docs/specifications/ctsum/>)

Inkorrekt:urn:cts:demo:sha[@kespeare.sonn-ets:35.1]@-

CTS URNs / Übung

- Sind diese URNs valide? Warum?

urn:cts:demo:goehte.faust.de:35.1

yes

Urn:cts:demo:goehte.faust.de:35.1

no

urn:cts:Demo:goehte.faust.de:35.1

yes

urn:cts:demo:goehte.faust.de:invalid35.1

yes

urn:cts:demo:goehte.faust.de:not-valid35.1

yes

urn:cts:demo:goehte.faust1-2.de:valid35.1

urn:cts:demo:shakespeare.sonnets:35.1@grieved-35.5@faults[1]

35.5@vaults[1] == 35.5@vaults

urn:cts:, namespace and textgroup benötigt, alles weitere optional

Groß- und kleinschreibung entscheidend(latinlit vs. LatinLit)

Beschreibend Teile können alle erlaubten Zeichen enthalten, auch Zahlen

Korrekt:urn:cts:demo:shakespeare.sonnets12.2:preface.line1

chapter2.line4 kann vorkommen vorchapter2.line1

Funktionstragende Zeichen dürfen nicht beschreibend verwendet werden

. :- @ []

(complete list <http://www.homermultitext.org/hmt-docs/specifications/ctsum/>)

Inkorrekt:urn:cts:demo:sha[@kespeare.sonn-ets:35.1]@-

CTS URNs / Übung

- Sind diese URNs valide? Warum?

urn:cts:demo:goehte.faust.de:35.1

yes

Urn:cts:demo:goehte.faust.de:35.1

no

urn:cts:Demo:goehte.faust.de:35.1

yes

urn:cts:demo:goehte.faust.de:invalid35.1

yes

urn:cts:demo:goehte.faust.de:not-valid35.1

yes

urn:cts:demo:goehte.faust1_2.de:valid35.1

no

urn:cts:goehte.faust1.de:valid35.1

urn:cts:demo:shakespeare.sonnets:35.1@grieved-35.5@faults[1]

35.5@vaults[1] == 35.5@vaults

urn:cts:, namespace and textgroup benötigt, alles weitere optional

Groß- und kleinschreibung entscheidend(latinlit vs. LatinLit)

Beschreibend Teile können alle erlaubten Zeichen enthalten, auch Zahlen

Korrekt:urn:cts:demo:shakespeare.sonnets12.2:preface.line1

chapter2.line4 kann vorkommen vorchapter2.line1

Funktionstragende Zeichen dürfen nicht beschreibend verwendet werden

. :- @ []

(complete list <http://www.homermultitext.org/hmt-docs/specifications/ctsum/>)

Inkorrekt:urn:cts:demo:sha[@kespeare.sonn-ets:35.1]@-

CTS URNs / Übung

- Sind diese URNs valide? Warum?

urn:cts:demo:goehte.faust.de:35.1

yes

Urn:cts:demo:goehte.faust.de:35.1

no

urn:cts:Demo:goehte.faust.de:35.1

yes

urn:cts:demo:goehte.faust.de:invalid35.1

yes

urn:cts:demo:goehte.faust.de:not-valid35.1

yes

urn:cts:demo:goehte.faust1_2.de:valid35.1

no

urn:cts:goehte.faust1.de:valid35.1

no

urn:cts:demo:goehte.faust.de:

urn:cts:demo:shakespeare.sonnets:35.1@grieved-35.5@faults[1]

35.5@vaults[1] == 35.5@vaults

urn:cts:, namespace and textgroup benötigt, alles weitere optional

Groß- und kleinschreibung entscheidend(latinlit vs. LatinLit)

Beschreibend Teile können alle erlaubten Zeichen enthalten, auch Zahlen

Korrekt:urn:cts:demo:shakespeare.sonnets12.2:preface.line1

chapter2.line4 kann vorkommen vorchapter2.line1

Funktionstragende Zeichen dürfen nicht beschreibend verwendet werden

. :- @ []

(complete list <http://www.homermultitext.org/hmt-docs/specifications/ctsum/>)

Inkorrekt:urn:cts:demo:sha[@kespeare.sonn-ets:35.1]@-

CTS URNs / Übung

- Sind diese URNs valide? Warum?

urn:cts:demo:goehte.faust.de:35.1

yes

Urn:cts:demo:goehte.faust.de:35.1

no

urn:cts:Demo:goehte.faust.de:35.1

yes

urn:cts:demo:goehte.faust.de:invalid35.1

yes

urn:cts:demo:goehte.faust.de:not-valid35.1

yes

urn:cts:demo:goehte.faust1_2.de:valid35.1

no

urn:cts:goehte.faust1.de:valid35.1

no

urn:cts:demo:goehte.faust.de:

yes

urn:cts:demo:shakespeare.sonnets:35.1@grieved-35.5@faults[1]

35.5@vaults[1] == 35.5@vaults

urn:cts:, namespace and textgroup benötigt, alles weitere optional

Groß- und kleinschreibung entscheidend(latinlit vs. LatinLit)

Beschreibend Teile können alle erlaubten Zeichen enthalten, auch Zahlen

Korrekt:urn:cts:demo:shakespeare.sonnets12.2:preface.line1

chapter2.line4 kann vorkommen vorchapter2.line1

Funktionstragende Zeichen dürfen nicht beschreibend verwendet werden

. :- @ []

(complete list <http://www.homermultitext.org/hmt-docs/specifications/ctsum/>)

Inkorrekt:urn:cts:demo:sha[@kespeare.sonn-ets:35.1]@-

CTS Funktionen

GetCapabilities ()

GetValidReff (urn, level)

GetFirstUrn (urn)

GetPrevNextUrn (urn)

GetLabel (urn)

GetPassage (urn)

GetPassagePlus (urn)

CTS Funktionen

GetCapabilities ()

GetValidReff (urn, level)

GetFirstUrn (urn)

GetPrevNextUrn (urn)

GetLabel (urn)

GetPassage (urn)

GetPassagePlus (urn)

Funktionsname und Parameter als GET in HTTP-Request

CTS Funktionen

GetCapabilities ()

Textinventory

GetValidReff (urn, level)

GetFirstUrn (urn)

GetPrevNextUrn (urn)

GetLabel (urn)

GetPassage (urn)

GetPassagePlus (urn)

```
<textgroup projid="greekLit:tlg0019" urn="urn:cts:greekLit:tlg0019">
  <groupname xml:lang="eng">Aristophanes</groupname>
  - <work projid="greekLit:tlg003" urn="urn:cts:greekLit:tlg0019.tlg003" xml:lan
    <title xml:lang="eng">Clouds</title>
    - <edition projid="greekLit:perseus-grc2">
      <label xml:lang="eng">Clouds</label>
      - <description xml:lang="eng">
        Perseus:bib:oclc,10582150, Aristophanes. Aristophanes Comoediae, ed. F
        and W.M. Geldart, vol. 2. F.W. Hall and W.M. Geldart. Oxford. Clarendon
        Oxford. 1907.
      </description>
      - <online docname="/db/repository/greekLit/tlg0019/tlg003/tlg0019.tlg003.p
        grc2.xml">
        <validate schema="teia.xsd"/>
        <namespaceMapping abbreviation="tei" nsURI="http://www.tei-c.org/
      - <citationMapping>
        <citation label="line" xpath="//tei:l[@n=?]" scope="/tei:TEI/tei:text
          /tei:body"/>
```

CTS Funktionen

GetCapabilities ()

GetValidReff (urn, level)

Alle „Kinder-URNs“ mit Tiefe [level] → Alle URNs für Kapitel 8

GetFirstUrn (urn)

GetPrevNextUrn (urn)

GetLabel (urn)

GetPassage (urn)

GetPassagePlus (urn)

```
<cts:GetValidReff>
- <cts:request>
  <cts:requestName>GetValidReff</cts:requestName>
  <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1</cts:urn>
  <cts:level>4</cts:level>
</cts:request>
- <cts:reply>
  - <cts:reff>
    <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1.</cts:urn>
    <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1</cts:urn>
    <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.pr</cts:urn>
    <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.pr.1</cts:urn>
    <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.pr.2</cts:urn>
```

CTS Funktionen

GetCapabilities ()

GetValidReff (urn, level)

GetFirstUrn (urn)

Erste „KindURN“ → 1st URN für Kapitel 8

GetPrevNextUrn (urn)

GetLabel (urn)

GetPassage (urn)

GetPassagePlus (urn)

```
<cts:GetFirstUrn>  
- <cts:request>  
  <cts:requestName>GetFirstUrn</cts:requestName>  
  <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:</cts:urn>  
</cts:request>  
- <cts:reply>  
  <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1</cts:urn>  
</cts:reply>  
</cts:GetFirstUrn>
```

CTS Funktionen

GetCapabilities ()

GetValidReff (urn, level)

GetFirstUrn (urn)

GetPrevNextUrn (urn)

Linke und rechte NachbarURNs → URNs links und rechts der URN für Zeile 8

GetLabel (urn)

GetPassage (urn)

GetPassagePlus (urn)

```
<cts:GetPrevNextUrn>
- <cts:request>
  <cts:requestName>GetPrevNextUrn</cts:requestName>
  <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.2.1</cts:urn>
  <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.2.1</cts:urn>
</cts:request>
- <cts:reply>
  - <cts:prevnext>
    - <cts:prev>
      <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.2</cts:urn>
    </cts:prev>
    - <cts:next>
      <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.2.2</cts:urn>
    </cts:next>
  </cts:prevnext>
</cts:reply>
</cts:GetPrevNextUrn>
```

CTS Funktionen

GetCapabilities ()

GetValidReff (urn, level)

GetFirstUrn (urn)

GetPrevNextUrn (urn)

GetLabel (urn)

Informelle Beschreibung der [urn] → „Shakespears Sonnet 36 Line 9“

```
Get <GetLabel>
  <request>
    <requestName>GetLabel</requestName>
    <requestUrn>urn:cts:pbcc:deu.luther1912:1.2.10-2.2.3</requestUrn>
  </request>
  - <reply>
    - <label>
      "Die Bibel in Deutsch. Luther von 1912. The Bible in German" from book "1", chapter "2", sentence "10" to book "2", chapter "2", sentence "3"
    </label>
  </reply>
</GetLabel>
```

CTS Funktionen

GetCapabilities ()

GetValidReff (urn, level)

GetFirstUrn (urn)

GetPrevNextUrn (urn)

GetLabel (urn)

GetPassage (urn)

Textabschnitt für eine [urn] → Der Text

GetPassagePlus (urn)

```
<cts:GetPassage>
- <cts:request>
  <cts:requestName>GetPassage</cts:requestName>
  <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.2.1</cts:urn>
</cts:request>
- <cts:reply>
  <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.2.1</cts:urn>
  - <cts:passage>
    Sacerdos uestra adhuc in lupanari uiueret nisi hominem occidi
    lenones, absint meretrices, ne quid parum sanctum occurrat, d
    occidit; at hercule lenonem non occidisti. Deducta es in lupan
    obscenum lectulum uocas ? de pudicitia sacerdotis hic quaerit
    mihi sacerdotem cuius precaria est castitas? Cum ex illo lupar
    FVLVI Sparsi.
  </cts:passage>
</cts:reply>
</cts:GetPassage>
```

CTS Funktionen

GetCapabilities ()

GetValidReff (urn, level)

GetFirstUrn (urn)

GetPrevNextUrn (urn)

GetLabel (urn)

GetPassage (urn)

GetPassagePlus (urn)

Alles zusammen ohne GetCapabilities

```
<cts:GetPassagePlus>
- <cts:request>
  <cts:requestName>GetPassagePlus</cts:requestName>
  <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.2.1</cts:urn>
</cts:request>
- <cts:reply>
  <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.2.1</cts:urn>
  - <cts:passage>
    Sacerdos uestra adhuc in lupanari uiueret nisi hominem occidisset.
    lenones, absint meretrices, ne quid parum sanctum occurrat, dum se
    occidit; at hercule lenonem non occidisti. Deducta es in lupanar, ac
    obscenum lectulum uocas ? de pudicitia sacerdotis hic quaeritur. "N
    mihi sacerdotem cuius precaria est castitas? Cum ex illo lupanari cr
    FVLVI Sparsi.
  </cts:passage>
  <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.2.1</cts:urn>
  <cts:groupname>phi1014</cts:groupname>
  <cts:label>Edition: Controversiae</cts:label>
  - <cts:prev>
    <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.2</cts:urn>
  </cts:prev>
  - <cts:next>
    <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.2.2</cts:urn>
  </cts:next>
  <cts:validreff/>
</cts:reply>
</cts:GetPassagePlus>
```

CTS Funktionen / Übung

GetCapabilities ()
GetPrevNextUrn (urn)

GetValidReff (urn, level)
GetLabel (urn)
(GetPassagePlus (urn))

GetFirstUrn (urn)
GetPassage (urn)

- Über welche Funktion findet man heraus, wie viele Editionen vom CTS verwaltet werden?

CTS Funktionen / Übung

GetCapabilities ()
GetPrevNextUrn (urn)

GetValidReff (urn, level)
GetLabel (urn)
(GetPassagePlus (urn))

GetFirstUrn (urn)
GetPassage (urn)

- Über welche Funktion findet man heraus, wie viele Editionen vom CTS verwaltet werden?
→ GetCapabilities
- Welche Funktion kann dazu genutzt werden, den gesamten Text einer Edition abzufragen?

CTS Funktionen / Übung

GetCapabilities ()
GetPrevNextUrn (urn)

GetValidReff (urn, level)
GetLabel (urn)
(GetPassagePlus (urn))

GetFirstUrn (urn)
GetPassage (urn)

- Über welche Funktion findet man heraus, wie viele Editionen vom CTS verwaltet werden?
→ GetCapabilities
- Welche Funktion kann dazu genutzt werden, den gesamten Text einer Edition abzufragen?
→ GetPassage

CTS Funktionen / Übung

GetCapabilities ()
GetPrevNextUrn (urn)

GetValidReff (urn, level)
GetLabel (urn)
(GetPassagePlus (urn))

GetFirstUrn (urn)
GetPassage (urn)

- Über welche Funktion findet man heraus, wie viele Editionen vom CTS verwaltet werden?
→ GetCapabilities
- Welche Funktion kann dazu genutzt werden, den gesamten Text einer Edition abzufragen?
→ GetPassage
- Welche Funktion ist nicht im Ergebnis von GetPassagePlus enthalten ?

CTS Funktionen / Übung

GetCapabilities ()
GetPrevNextUrn (urn)

GetValidReff (urn, level)
GetLabel (urn)
(GetPassagePlus (urn))

GetFirstUrn (urn)
GetPassage (urn)

- Über welche Funktion findet man heraus, wie viele Editionen vom CTS verwaltet werden?
→ GetCapabilities
- Welche Funktion kann dazu genutzt werden, den gesamten Text einer Edition abzufragen?
→ GetPassage
- Welche Funktion ist nicht im Ergebnis von GetPassagePlus enthalten(urn) ?
→ GetCapabilities
- Was müsste man tun, um alle CTS URNs abzufragen (innere & äußere Hierarchie) abzufragen?

CTS Funktionen / Übung

GetCapabilities ()
GetPrevNextUrn (urn)

GetValidReff (urn, level)
GetLabel (urn)
(GetPassagePlus (urn))

GetFirstUrn (urn)
GetPassage (urn)

- Über welche Funktion findet man heraus, wie viele Editionen vom CTS verwaltet werden?
→ GetCapabilities
- Welche Funktion kann dazu genutzt werden, den gesamten Text einer Edition abzufragen?
→ GetPassage
- Welche Funktion ist nicht im Ergebnis von GetPassagePlus enthalten(urn) ?
→ GetCapabilities
- Was müsste man tun, um alle CTS URNs abzufragen (innere & äußere Hierarchie) abzufragen?
→ 1 GetCapabilities → foreach edition GetValidReff (urn, 10000000)
→ 2 GetCapabilities → foreach edition GetPrevNext (urn) until no response
- Welcher Weg wäre wahrscheinlich am effizientesten?

CTS Funktionen / Übung

GetCapabilities ()
GetPrevNextUrn (urn)

GetValidReff (urn, level)
GetLabel (urn)
(GetPassagePlus (urn))

GetFirstUrn (urn)
GetPassage (urn)

- Über welche Funktion findet man heraus, wie viele Editionen vom CTS verwaltet werden?
→ GetCapabilities
- Welche Funktion kann dazu genutzt werden, den gesamten Text einer Edition abzufragen?
→ GetPassage
- Welche Funktion ist nicht im Ergebnis von GetPassagePlus enthalten(urn) ?
→ GetCapabilities
- Was müsste man tun, um alle CTS URNs abzufragen (innere & äußere Hierarchie) abzufragen?
→ 1 GetCapabilities → foreach edition GetValidReff (urn, 10000000)
→ 2 GetCapabilities → foreach edition GetPrevNext (urn) until no response

CTS Requests

[http://cts.informatik.uni-leipzig.de/pbc/cts/?request=GetPassage&urn=urn:cts:pbcbible.parallel.eng.kingjames:1.3.5-1.3.8@the\[2\]](http://cts.informatik.uni-leipzig.de/pbc/cts/?request=GetPassage&urn=urn:cts:pbcbible.parallel.eng.kingjames:1.3.5-1.3.8@the[2])

CTS Requests

[http://cts.informatik.uni-leipzig.de/pbc/cts/?request=GetPassage&urn=urn:cts:pbcbible.parallel.eng.kingjames:1.3.5-1.3.8@the\[2\]](http://cts.informatik.uni-leipzig.de/pbc/cts/?request=GetPassage&urn=urn:cts:pbcbible.parallel.eng.kingjames:1.3.5-1.3.8@the[2])

Server, Endpoint, Access Point

CTS Requests

[http://cts.informatik.uni-leipzig.de/pbc/cts/?request=GetPassage&urn=urn:cts:pbcbible.parallel.eng.kingjames:1.3.5-1.3.8@the\[2\]](http://cts.informatik.uni-leipzig.de/pbc/cts/?request=GetPassage&urn=urn:cts:pbcbible.parallel.eng.kingjames:1.3.5-1.3.8@the[2])

Funktionsname

GetPassage, GetPassagePlus, GetLabel,
GetCapabilities, GetValidReff, GetFirstUrn,
GetPrevNextUrn

CTS Requests

[http://cts.informatik.uni-leipzig.de/pbc/cts/?request=GetPassage&urn=urn:cts:pbc:bible.parallel.eng.kingjames:1.3.5-1.3.8@the\[2\]](http://cts.informatik.uni-leipzig.de/pbc/cts/?request=GetPassage&urn=urn:cts:pbc:bible.parallel.eng.kingjames:1.3.5-1.3.8@the[2])

Funktionsparameter

URN, (Zitation)level,...

CTS Response

[http://cts.informatik.uni-leipzig.de/pbc/cts/?request=GetPassage&urn=urn:cts:pbcbible.parallel.eng.kingjames:1.3.5-1.3.8@the\[2\]](http://cts.informatik.uni-leipzig.de/pbc/cts/?request=GetPassage&urn=urn:cts:pbcbible.parallel.eng.kingjames:1.3.5-1.3.8@the[2])

```
<GetPassage>
  <request>
    <requestName>GetPassage</requestName>
    <requestUrn>urn:cts:pbcbible.parallel.eng.kingjames:1.3.5-1.3.8@the[2]</requestUrn>
  </request>
  <reply>
    <urn>urn:cts:pbcbible.parallel.eng.kingjames:1.3.5-1.3.8@the[2]</urn>
    <passage>For God doth know that in the day ye eat thereof , then your eyes shall be opened , and ye shall be as gods , knowing good and evil . And when the woman saw that the tree was good for food , and that it was pleasant to the eyes , and a tree to be desired to make one wise , she took of the fruit thereof , and did eat , and gave also unto her husband with her ; and he did eat . And the eyes of them both were opened , and they knew that they were naked ; and they sewed fig leaves together , and made themselves aprons . And they heard the</passage>
    <license>Public Domain</license>
    <source>Retrieved via Canonical Text Service http://cts.informatik.uni-leipzig.de/pbc/cts/ with CTS URN urn:cts:pbcbible.parallel.eng.kingjames:1.3.5-1.3.8@the[2]</source>
  </reply>
</GetPassage>
```

CTS Beispielabfragen

[GetCapabilities](#)

The text inventory with every CTS URN on document level and the corresponding meta information.

[GetPassage](#)

The text passage that is referenced by a given CTS URN.

[GetLabel](#)

An informal description for the text passage that is referenced by a given CTS URN. For example useful for tooltips.

[GetValidReff](#)

All the CTS URNs that "belong to" a given CTS URN on a given citation level. For example, every *sentence CTS URN* that "belongs to" a given *chapter CTS URN*.

[GetPrevNextUrn](#)

The previous and next static CTS URNs for a given CTS URN in document order.

[GetFirstUrn](#)

The first CTS URN that "belongs to" a given CTS URN. For example, the first *sentence CTS URN* that "belongs to" a given *chapter CTS URN*.

[GetPassagePlus](#)

The combined information for a given CTS URN.

Lizenz Handling

```
-<edition urn="urn:cts:dta:343012.de.transcript:">
  -<title>
    Christliche Leich- vnd Ehren-Predigt bey der Sepultur der weiland Erbar[n] Ehr vnd Thugendreichen Frawen Barbara Pottermannin
  </title>
  <author>Seidemannus Martinus</author>
  -<license>
    Dieses Werk steht unter der „Creative Commons Namensnennung-Weitergabe unter gleichen Bedingungen 3.0 Deutschland Lizenz“ (CC BY-SA)
  </license>
  -<source>
    http://www.deutschestextarchiv.de/343012 http://www.deutschestextarchiv.de/book/download_xml/343012 http://www.deutschestextarchiv.de/b
    http://www.deutschestextarchiv.de/book/download_text/343012 343012 200004 urn:nbn:de:kobv:b4-200004-1 http://www.deutschestextarchiv
    /book/download_xml/343012 http://www.deutschestextarchiv.de/book/download_html/343012 http://www.deutschestextarchiv.de/book/downlo
    urn:nbn:de:kobv:b4-200004-1 Universitätsbibliothek Breslau, 4 A 277/7 / 343012 Universitätsbibliothek Breslau, 4 A 277/7 / 343012 retrieved
  </source>
  <publicationDate>1625</publicationDate>
  <contentType>xml</contentType>
</edition>
```

Lizenzen wie Creative Commons CC-BY erfordern Nennung von Lizenz & Quelle bei jedem Zitat

Bei CTS eigentlich nicht vorgesehen

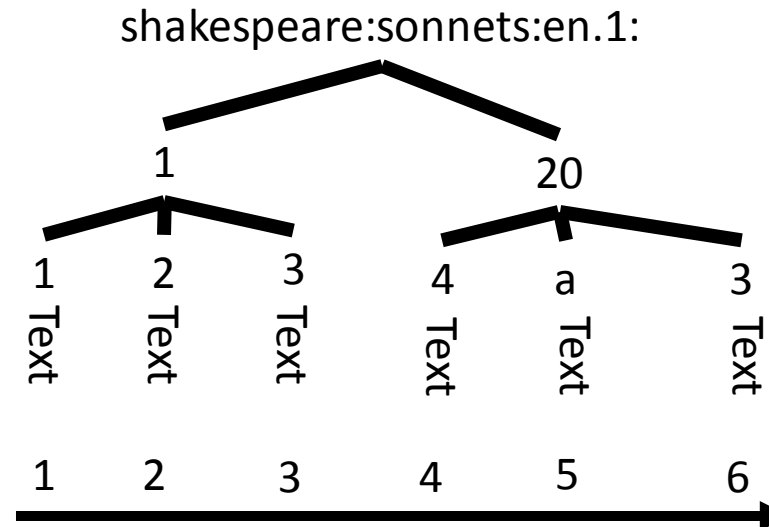
In Zusammenarbeit mit BBAW für die ASV-Implementierung von CTS entwickelt (unique)

```
-<GetPassage>
  -<request>
    <requestName>GetPassage</requestName>
    <requestUrn>urn:cts:dta:343012.de.transcript:1-2</requestUrn>
  </request>
  -<reply>
    <urn>urn:cts:dta:343012.de.transcript:1-2</urn>
    -<passage>
      Chriftliche Leich- vnd Ehren-Predigt bey der Sepultur der weiland I
      Handelsmans/ vnd wolverordneten Königl. Hoff- Schöppens in Lev
      die Clofter Kirche mit Chriftlichen Ceremonien befattet worden; G
    </passage>
    -<license>
      Dieses Werk steht unter der „Creative Commons Namensnennung-1
    </license>
    -<source>
      http://www.deutschestextarchiv.de/343012 http://www.deutsche
      /book/download_text/343012 343012 200004 urn:nbn:de:kobv:b4
      http://www.deutschestextarchiv.de/book/download_html/343012 h
      4 A 277/7 / 343012 Universitätsbibliothek Breslau, 4 A 277/7 / 34
    </source>
  </reply>
</GetPassage>
```

Mapping Hierarchische IDs auf Text

- :1
- :1.1
- :1.1.1 O Tannenbaum, O Tannenbaum,
- :1.1.2 Wie treu sind deine Blätter.
- :1.1.3 Du grünst nicht nur zur Sommerzeit,
- :1.1.4 Nein auch im Winter wenn es schneit.
- :1.1.5 O Tannenbaum, O Tannenbaum,
- :1.1.6 Wie grün sind deine Blätter!
- :1.2
- :1.2.1 O Tannenbaum, O Tannenbaum,
- :1.2.2 Du kannst mir sehr gefallen!
- :1.2.3 Wie oft hat schon zur Winterszeit
- :1.2.4 Ein Baum von dir mich hoch erfreut!
- :1.2.5 O Tannenbaum, O Tannenbaum,
- :1.2.6 Du kannst mir sehr gefallen!
- :1.3
- :1.3.1 O Tannenbaum, O Tannenbaum,
- :1.3.2 Dein Kleid will mich was lehren:
- :1.3.3 Die Hoffnung und Beständigkeit
- :1.3.4 Gibt Mut und Kraft zu jeder Zeit!
- :1.3.5 O Tannenbaum, O Tannenbaum,
- :1.3.6 Dein Kleid will mich was lehren.

Flexibel genug, um für jeden
Text anwendbar zu sein
Persistent



Div-View

urn:cts:songs:christmas.ohtennenbaum.de.1:1-1.2.4

<passage>

O Tannenbaum, O Tannenbaum, (...) Wie grün sind deine Blätter! O Tannenbaum, O Tannenbaum, (...) Ein Baum von dir mich hoch erfreut!

</passage>

<passage>

<div1 n="1" type="song">

<div2 n="1" type="strophe">

<div3 n="1" type="line">O Tannenbaum, O Tannenbaum, </div3>

(...)

<div3 n="6" type="line">Wie grün sind deine Blätter! </div3>

</div2>

<div2 n="2" type="strophe">

<div3 n="1" type="line">O Tannenbaum, O Tannenbaum, </div3>

(...)

<div3 n="6" type="line">Ein Baum von dir mich hoch erfreut!</div3>

</div2>

</div1>

</passage>

Generic Reader

demo - dta - greekLit - latinLit - songs

3 - tlg0004 - tlg0005 - tlg0006 - tlg0007 - tlg0008 - tlg0010 - tlg0011 - tlg0012 - tlg0013 - tlg0014 - tlg0016 - tlg0017 - tlg0019

tlg001 - tlg002 - tlg003 - tlg004 - tlg005 - tlg006 - tlg007 - tlg008

perseus_eng1 - perseus_eng2 - perseus_grc1 - perseus_grc2

View Mode: XML Styled Set Citation Export Citation

Sophocles
Trachiniae (English) Machine readable text Tags

Hyllus

I will go, Mother. Had I known the substance of these prophecies, I would have long been at his side. As it was, my father's usual good fortune did not allow me to fear for him, nor to be overly anxious. Now that I have the knowledge, I will spare no pains to learn the whole truth in this matter.

Deianeira

Go, then, my son. For prosperity yields advantage even for him who learns of it late.
(Exeunt Hyllus, on one side, and the Nurse into the house.)
(Enter the Chorus on the other side.)

Chorus

You, to whom Night gives birth when she is vanquished and despoiled of her starry crown, and whom, as you blaze, she
prize, now like some mournful bird, never rests her eyes' longing
the might be without tears, but nourishing a well-remembered fear for her husband's travels she is constantly
afflict by her anxious, widowed marriage-bed and in her misery anticipates misfortune.

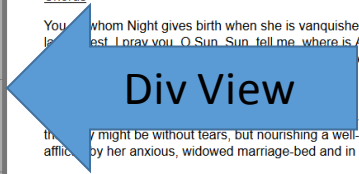
Chorus

For just as one may see billow after billow advancing and passing over the wide deep before the tireless south-wind, or the north, so the great toil of his life, stormy as the Cretan sea, now whirls back the heir of Cadmus, now exalts him. But some god always keeps him unerring from the house of Hades.
(To Deianeira.)

Chorus

With respect I reproach you for your weeping, but still I will speak in directness. You must not, I say, wear away fair boys.

Copyright 2014 Leipzig University // Martin Reckziegel



demo - dta - greekLit - latinLit - songs

3 - tlg0004 - tlg0005 - tlg0006 - tlg0007 - tlg0008 - tlg0010 - tlg0011 - tlg0012 - tlg0013 - tlg0014 - tlg0016 - tlg0017 - tlg0019

tlg001 - tlg002 - tlg003 - tlg004 - tlg005 - tlg006 - tlg007 - tlg008

perseus_eng1 - perseus_eng2 - perseus_grc1 - perseus_grc2

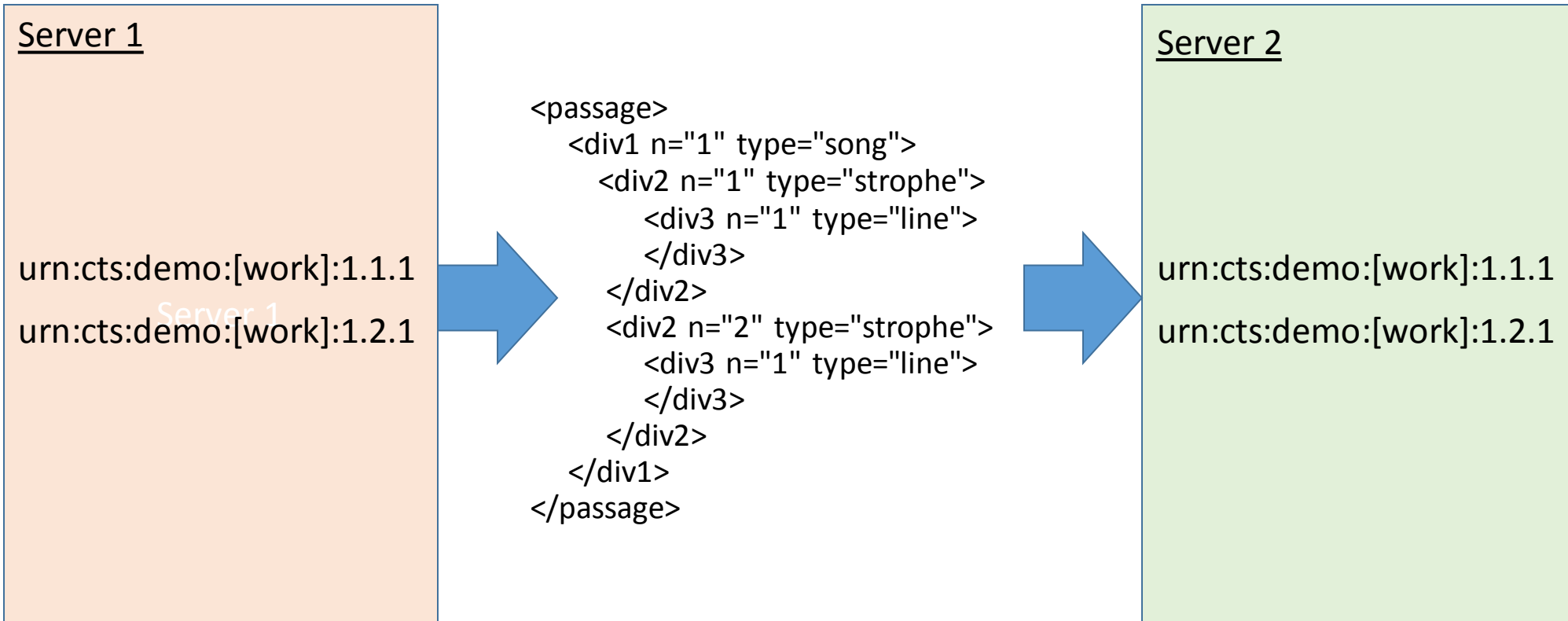
View Mode: XML Styled Set Citation Export Citation

Sophocles
Trachiniae (English) Machine readable text Tags

```
</milestone>
  if he finds safety, or we perish along with him.
</p>
</sp>
<sp TEIform="sp">
  <speaker TEIform="speaker">
    | Hyllus
  </speaker>
  <p TEIform="p">
    I will go, Mother. Had I known the substance of these prophecies, I would have long been at his side. As it
    was, my father's usual good fortune did not allow me to fear for him, nor to be overly anxious.
    <milestone ed="p" n="90" unit="line" TEIform="milestone">
    </milestone>
    Now that I have the knowledge, I will spare no pains to learn the whole truth in this matter.
  </p>
</sp>
<sp TEIform="sp">
  <speaker TEIform="speaker">
    Deianeira
  </speaker>
  <p TEIform="p">
    Go, then, my son. For prosperity yields advantage even for him who learns of it late.
    <stage TEIform="stage">
    | Exeunt Hyllus, on one side, and the Nurse into the house.
    </stage>
  </p>
</sp>
<milestone unit="choral" TEIform="milestone">
</milestone>
<milestone unit="strophe" n="1" TEIform="milestone">
</milestone>
<stage TEIform="stage">
  | Enter the Chorus on the other side.
</stage>
<sp TEIform="sp">
  <speaker TEIform="speaker">
    Chorus
  </speaker>
  <p TEIform="p">
    You, to whom Night gives birth when she is vanquished and despoiled of her starry crown,
    <milestone ed="p" n="95" unit="line" TEIform="milestone">
    </milestone>
    and whom, as you blaze, she lays to rest. I pray you, O Sun, Sun, tell me, where is Alcmena's son, where
    dwells her child? O Shining god with your bright flash,
    <milestone ed="p" n="100" unit="line" TEIform="milestone">
    </milestone>
  </p>
</sp>
```

Copyright 2014 Leipzig University // Martin Reckziegel

CTS Cloning



Datenbestand CTS

Datensatz	Dokumente	Statische Referenzen (CTS URNs)	HDD Größe Index/Daten (MB)	Sprache (n)
Textgrid	91133	7284050	2730/2621	Deu
Deutsches Text Archiv	8190	21911559	7841/12887	Deu
Perseus	1693	273126	490/65	Grc, Lat, Farsi, Eng
Parallel Bible Corpus	831	8818067	3330/4972	Ca 800
TED Talk Subtitle Transripts	52987	15292408	2642/2507	105

Weitere Datensätze:

Shakespeare, Briefe Berliner-Intellektuelle um 1800, Croatiae Auctores Latini,
German Political Speeches, Ali's monthly journal al-Muqtabas, Fragmentary Texts, ...

CLARIN

- CLARIN
 - European Research Infrastructure for Language Resources and Technology
- Virtual Language Observatory
 - Tool für facettierte Suche

The screenshot displays the CLARIN Virtual Language Observatory search results page. At the top, the title "CLARIN Virtual Language Observatory" is followed by a brief instruction: "Use the search bar to start searching the VLO, or continue to browse everything and use facets to narrow down to your area of interest or discover new resources." Below this is a search bar and a "See all records" button. The main content area shows a list of search results, each with a title, a description, and a set of icons representing different facets. The results are as follows:

Collection	Description	Facets
ukr_newscrawl_2011_1M	1.000.000 sentences of a Ukrainian newspaper corpus based on material crawled in 2011	Language, Collection, Resource type, Modality, Format, Keyword, Availability, Search options
rus_news_2013_1M	1.000.000 sentences of a Russian newspaper corpus based on material from 2013	Language, Collection, Resource type, Modality, Format, Keyword, Availability, Search options
deu_news_2012_3M	3.000.000 sentences of a German newspaper corpus based on material from 2012	Language, Collection, Resource type, Modality, Format, Keyword, Availability, Search options
deu_news_2010_10K	10.000 sentences of a German newspaper corpus based on material from 2010	Language, Collection, Resource type, Modality, Format, Keyword, Availability, Search options
nld_news_2012_300K	300.000 sentences of a Dutch newspaper corpus based on material from 2012	Language, Collection, Resource type, Modality, Format, Keyword, Availability, Search options
spa_news_2011_3M	3.000.000 sentences of a Spanish newspaper corpus based on material from 2011	Language, Collection, Resource type, Modality, Format, Keyword, Availability, Search options
eng_news_2013_1M	1.000.000 sentences of an English newspaper corpus based on material from 2013	Language, Collection, Resource type, Modality, Format, Keyword, Availability, Search options
ces_news_2012_1M	1.000.000 sentences of a Czech newspaper corpus based on material from 2012	Language, Collection, Resource type, Modality, Format, Keyword, Availability, Search options

Integration in CLARIN

- Presentation in VLO:

urn:cts:pbcbible

urn:cts:pbcbible.parallel.arb.norm:

urn:cts:pbcbible.parallel.ceb.bugna:

urn:cts:pbcbible.parallel.ces.kralicka:

...



[Record details](#)

[Resources \(0\)](#)

[Availability](#)

[All met](#)

Use the tree below to explore the hierarchy this record is part of.

Parallel Bible Corpus Canonical Text Service

├ The Bible in Arabic

├ Cebuano Ang Biblia (Bugna Version)

└ Czech Bible Kralicka. Version of 1613



Record 3 of 21



The Bible in French. King James Version.

Show the original provider's page for this record



Record details Resources (68) Availability All metadata Technical details Hierarchy

Name	Type		
cts	other	...	>
kingjames.xml	other	...	>
cts	text document	...	>
cts	text document	...	>
cts	text document	...	>
cts	text document	...	>
cts	text document	...	>
cts	text document	...	>
cts	text document	...	>
cts	text document	...	>
cts	text document	...	>
cts	text document	...	>
cts	text document	...	>
cts	text document	...	>

Mime type: text/xml

Link: <http://cts.informatik.uni-leipzig.de/pbc/cts/?request=GetPassage&urn=urn:cts:pb: bible.parallel.fra.kingjames:10>

Record 3 of 21

< previous next >

The Bible in French. King James Version.

Show the original provider's page for this record

Record details Resources (68) Availability All metadata Technical details Hierarchy

Name	Type		
cts	other	...	>
kingjames.xml	other	...	>
cts → urn:cts:pbcbible:parallel.fra.kingjames:1	text document	...	>
cts → urn:cts:pbcbible:parallel.fra.kingjames:2	text document	...	>
cts → urn:cts:pbcbible:parallel.fra.kingjames:3	text document	...	>
cts → urn:cts:pbcbible:parallel.fra.kingjames:4	text document	...	>
cts → urn:cts:pbcbible:parallel.fra.kingjames:5	text document	...	>
cts → urn:cts:pbcbible:parallel.fra.kingjames:6	text document	...	>
cts → urn:cts:pbcbible:parallel.fra.kingjames:7	text document	...	>
cts → urn:cts:pbcbible:parallel.fra.kingjames:8	text document	...	>
cts	text document	...	>
cts	text document	...	>
cts	text document	...	>

Mime type: text/xml

Link: <http://cts.informatik.uni-leipzig.de/pbc/cts/?request=GetPassage&urn=urn:cts:pbcbible:parallel.fra.kingjames:10>

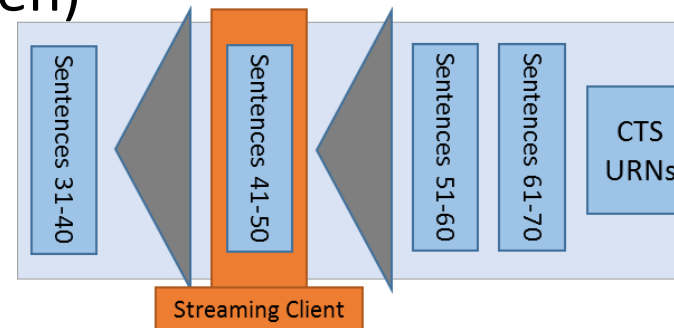
Positionierung von CTS in der Informatik

- Representational State Transfer (REST)
 - Client-server: Request basierter Webservice
 - Stateless: Keine Funktion verwendet Caching, jeder Request ist in sich abgeschlossen
 - Caching möglich: Persistenz impliziert Unveränderlichkeit der Antwort
 - Uniformes Interface: Anwendungs- und Systemunabhängigkeit per Definition
 - Schichtensystem: Nicht relevant, aber Implementierung arbeitet teilweise „schichtig“
 - Code-on-demand: Nicht relevant, außer für CTS URNs als Query Language
- CTS kann als RESTFul Webservice verstanden werden
 - Keine Eigenschaft von CTS widerspricht den Anforderungen eines RESTFul Webservice

Positionierung von CTS in der Informatik

- Nutzen

- Normalisierter Zugang zu Textdaten
 - Interoperabilität, Generalisierung, Uniformität
- Separater Zugang zu Strukturinformationen
 - Uniformer Zugang zu Strukturinformationen erlaubt Tools, die diese analysieren (CTRACER und Text Alignment)
- Flexible Granularität
- Text Streaming (Mobilanwendungen)



Positionierung von CTS in den (Digital) Humanities

- Anforderungsanalyse (Teilweise persönliches Feedback)
 - Persistente Zitierbarkeit: Per Definition
 - Flexible Granularität: Per Definition
 - Technologieunabhängigkeit: Per Definition
 - Lizenzhandling: Nicht Teil von CTS, trotzdem implementiert
 - Zugang / Einschränkung: Aufgabe des Serveradministrators
 - Bedienbarkeit: Admin Tool, Test Suite
 - Sprachunabhängigkeit: UTF-8 (konfigurierbar)

Positionierung von CTS in den (Digital) Humanities

- Nutzen
 - Archivierung
 - Dezentrales, selbstorganisiertes Archivsystem kombiniert mit zentralen Archiven
 - Interoperabilität zwischen Tools und Daten
 - Uniformes, anwendungs- und systemunabhängiges Dateninterface
 - Referenzieren von Textstellen
 - Publikation
 - Verfügbarkeit der Daten für Analyseworkflows

Positionierung von CTS mit Bezug zu Big Data

- Die 4 Big V's
 - Velocity -> Nur bei Indexerstellung relevant, also vernachlässigbar
Relevant für OCR, Zitatanalyse, aufwändige Workflows
 - Veracity -> Hier nicht relevant da korrigierte Texte
Relevant für bspw. OCR
 - Variety -> Heterogenität in Datenformaten & Zugriff
Siehe Interoperabilität
 - Volume -> Für CTS unkritisch (Siehe nächste Folie)

Positionierung von CTS mit Bezug zu Big Data

- Volume
 - Primärdaten (Text) typischerweise klein (<10 Gigabyte)
 - Sekundärdaten (Annotationen, Metadaten) können Volume beliebig aufblähen, aber nicht relevant für CTS
 - Geringes Primärdaten-Volume charakteristisch für Datentyp Text wegen Domainbezug der Dokumente eines Korpus
- CTS ist hauptsächlich im Bereich Variety zu positionieren

Positionierung von CTS mit Bezug zu Big Data

- Nutzen
 - CTS selbst eröffnet Parallelisierungsstrategien
 - Generische Parallelisierung mit flexibler, intuitiver Granularität (Kapitel, Dokument,...)
 - Verteiltes Textrepository
 - Zentrale Verwaltung der Namespaces, dezentrale Verteilung der Daten
 - *Sometimes Big Data means a lot of small data*

Kurze Pause...

- Nächste Woche geht es weiter mit technischen Details.