

# European Summer School 2017

Text Mining with Canonical Text Services  
Theory Session 9 – Interoperability

# Text Data Sources

Local “private” Files

Document Archives (Online/Offline)


Sourcecode Repositories (Github,...)

Forums, Social Media


\*Everything that can store text\*

# Selected Online Archives

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 Einstellun

- CONTACT

---

- DR.  
ALEXANDER GEYKEN  
Arbeitsstellenleiter Digitales Wörterbuch der  
deutschen Sprache  
Tel: +49 (0)30 20370 390  
→ [geyken@bbaw.de](mailto: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

# Selected Online Archives

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 with the text 'SPIEGEL ONLINE KULTUR' and a search icon. A secondary navigation bar lists various categories like Politik, Wirtschaft, Panorama, Sport, Kultur, etc. The main content area features a banner for 'Projekt Gutenberg-DE' with a search bar and a 'Suchen' button. Below the banner is a navigation menu with links for Startseite, Genres, Neue Texte, etc. The main text area displays the title 'William Shakespeare: Hamlet, Prinz von Dänemark. (Übersetzer Wieland) - Kapitel 2' and the beginning of the text, including the stage directions and the first dialogue between Bernardo and Francisco.

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

SPIEGEL ONLINE KULTUR

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/>

- Growing Collection of German Literature without Copyright restrictions
- >5000 Documents from >1100 Authors
- Community based proof reading [www.gaga.net](http://www.gaga.net)
- Text structure very flexible

# Selected Online Archives

**HATHI TRUST**  
Digital Library

**LOG IN** +

**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 from editions from various sources
- Pictures, PDF, no text (?)
- 30 Languages
- Login required for full access
- Selected documents as Public Domain
- Research, Teaching, Literature, Politics,...

# Selected Online Archives

The screenshot shows the homepage of the Wortschatz Portal. At the top, there is a logo for 'WORTSCHATZ UNIVERSITÄT LEIPZIG' and a search bar with the text 'Willkommen beim Wortschatz-Portal.' and 'Wort:'. Below the search bar is a 'Suche!' button and a checkbox for 'Beachte Groß-/Kleinschreibung'. 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 this, there are three main sections: 'Wörter des Tages', 'Neues Wortschatzportal', and 'Wortschatzportal (alt)'. At the bottom, there are three smaller sections: 'Wörterbuch', 'Webservices', and 'Feedback'.

**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 Portales

**Feedback**  
Feedback der Nutzer zu den verschiedenen Services der

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

- Word as Resource
- Cooccurrence, Trend Analysis, Example sentences,...
- Updated regularly based on selected Online Resources

# Selected Online Archives

The screenshot shows the Virtual Language Observatory (VLO) website. At the top, the VLO logo is displayed with the tagline "Explore the world of language resources and technology from different perspectives". To the right of the logo are several partner logos: CLARIN, WALS, THE CLARIN CENTER, ERA, DOBES, and others. Below the header, there is a navigation bar with "VLO > Faceted search" on the left and "Link | Report | Help" on the right. The main content area is titled "SEARCH" and features a search input field with a "Search" button. Below the search bar, it indicates "ALL RECORDS" and "780945 results". A pagination control shows "Showing 1 to 10" records, with the first page selected. The first record is titled "Persian MS 364" and includes an "Expand" link. The description for this record is: "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." Below the record, it shows "Resources: | 1 other |". The second record is titled "Islamic Arabic 1465" and also has an "Expand" link. The description for this record is: "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". On the right side of the search results, there is a "NARROW DOWN" sidebar with the instruction "Use the categories below to limit the search results to those matching the selected value(s)". The sidebar lists several categories: LANGUAGE, COLLECTION, RESOURCE TYPE, CONTINENT, COUNTRY, MODALITY, GENRE, and SUBJECT, each with a plus sign to its left.

Facetted Federated Search  
- Facets include various properties (genre, modality, organisation,...)  
- 780945 Documents

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

# Selected Online Archives

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

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

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

<http://historicaltextarchive.com/>

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

(...)



# Data Heterogeneity

No common format & access

No real distinction between text structure, meta information and text

<speech> vs <speaker> vs Elliot:

Files and tools incompatible accross projects

# Research Infrastructures

Environments for DH that allow for an easy and efficient

- evaluation
- selection
- reuse, and
- combination of NLP tools

in combination with an easy access to

- different kinds of textual data
- that have well assigned metadata

# CLARIN

- CLARIN
  - European Research Infrastructure for Language Resources and Technology
- Virtual Language Observatory
  - Tool for faceted distributed search

The screenshot shows the CLARIN Virtual Language Observatory search interface. At the top, the title "CLARIN Virtual Language Observatory" is displayed, 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." A "See all records" button is located below the instruction. The search bar is empty, and the results section shows "Showing all 922801 records" and "Results per page: 10". The results are displayed in a list format, with each entry including a language code, a collection name, a description of the corpus, and a set of icons for actions like download and share. The list includes entries for Ukrainian, Russian, German, Dutch, Spanish, English, and Czech newspaper corpora.

Language	Collection	Resource type	Modality	Format	Keyword	Availability	Search options
ukr	ukr_newscrawl_2011_1M	1,000,000 sentences of a Ukrainian newspaper corpus based on material crawled in 2011					
rus	rus_news_2013_1M	1,000,000 sentences of a Russian newspaper corpus based on material from 2013					
deu	deu_news_2012_3M	3,000,000 sentences of a German newspaper corpus based on material from 2012					
deu	deu_news_2010_10K	10,000 sentences of a German newspaper corpus based on material from 2010					
nld	nld_news_2012_300K	300,000 sentences of a Dutch newspaper corpus based on material from 2012					
spa	spa_news_2011_3M	3,000,000 sentences of a Spanish newspaper corpus based on material from 2011					
eng	eng_news_2013_1M	1,000,000 sentences of a English newspaper corpus based on material from 2013					
ces	ces_news_2012_1M	1,000,000 sentences of a Czech newspaper corpus based on material from 2012					

# CLARIN

provide access to a huge variety of resources (data & tools) via webservices

Metadata and Federated Search

workflows via chaining of webservices

(simple to use) webapplications, no local installations

“safe” citation based on PIDs

improvement of the distribution of resources and tools in various communities

# CLARIN's Virtual Language Observatory



The screenshot shows the CLARIN Virtual Language Observatory (VLO) search interface. The header features the VLO logo and a tagline: "Explore the world of language resources and technology from different perspectives". A navigation bar shows the current search path: "VLO > Faceted search > Search: 'Leipzig' > Selections: Written Corpus x German x". A search bar contains the text "Leipzig" and a "Search" button. The search results section displays "6 results" and "Showing 1 to 6". The first result is "Wortschatz", described as "Collected from newspaper texts, webcrawling, etc.: words (+frequency), cooccurrences (+graph), left/right neighbours, example sentences". Below it, a resource "deu\_news\_2008\_100K" is shown, described as "100.000 sentences of a German newspaper corpus based on material from 2008". A "NARROW DOWN" sidebar on the right allows filtering by LANGUAGE (German x), COLLECTION, RESOURCE TYPE (Written Corpus x), COUNTRY, GENRE, and FORMAT.

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

VLO > [Faceted search](#) > [Search: "Leipzig"](#) > [Selections: Written Corpus x German x](#) [Permalink](#) | [Report](#) | [Help](#)

SEARCH

Leipzig

**SEARCH RESULTS**

6 results Showing 1 to 6

**Wortschatz** Expand  
Collected from newspaper texts, webcrawling, etc.: words (+frequency), cooccurrences (+graph), left/right neighbours, example sentences  
Resources: | 1 other |

**deu\_news\_2008\_100K** Expand  
100.000 sentences of a German newspaper corpus based on material from 2008  
Resources: | 1 text document | 1 other |

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

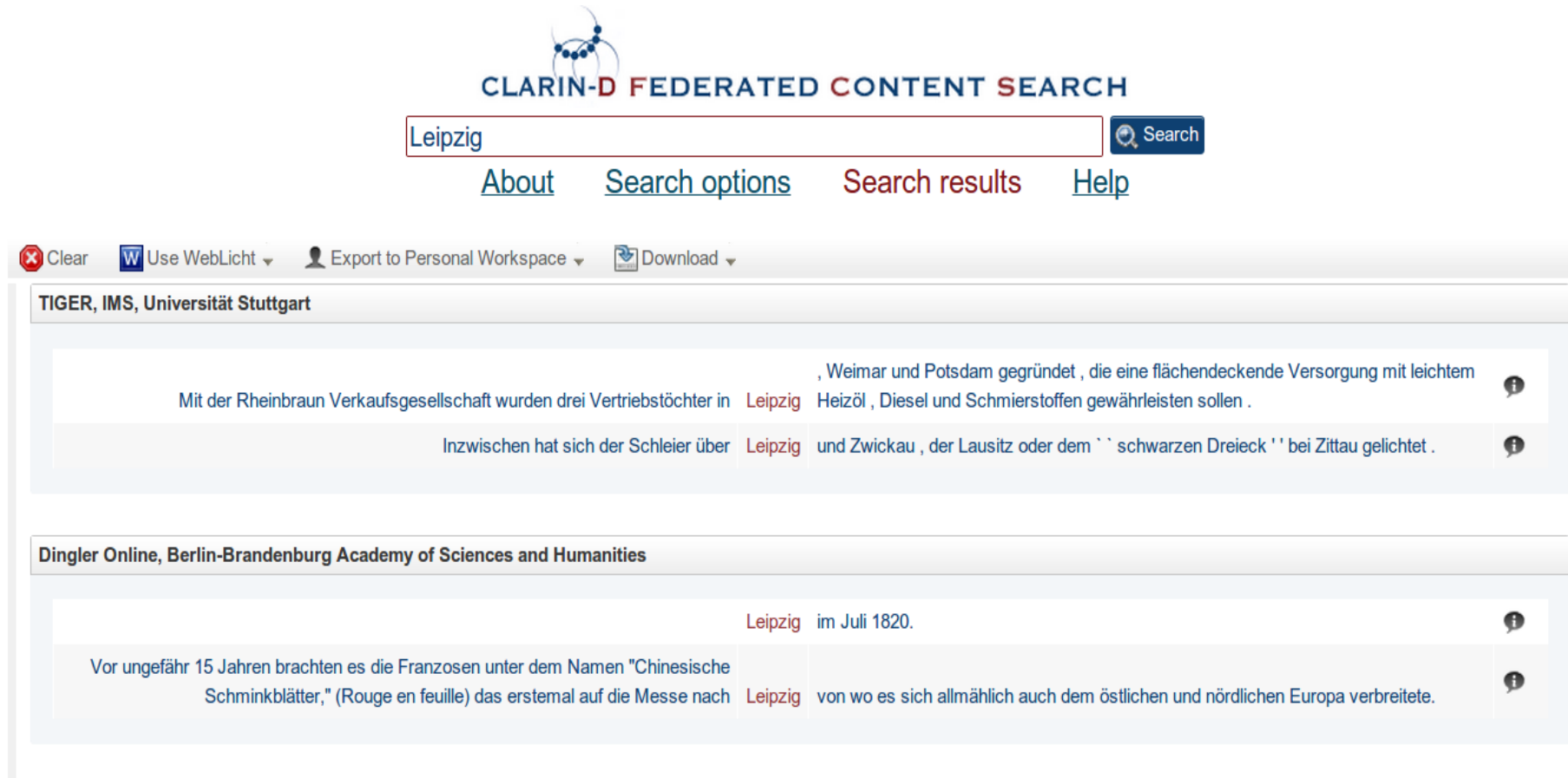
- + LANGUAGE
  - German x**
- + COLLECTION
- RESOURCE TYPE
  - Written Corpus x**
- + COUNTRY
- + GENRE
- + FORMAT

# CLARIN's Virtual Language Observatory

The screenshot displays the CLARIN-D Federated Content Search interface. At the top, the logo features a stylized network of blue nodes and lines. Below it, the text 'CLARIN-D FEDERATED CONTENT SEARCH' is centered. A search bar contains the text 'Leipzig', followed by a 'Search' button with a magnifying glass icon. Below the search bar are four navigation links: 'About', 'Search options', 'Search results' (highlighted in red), and 'Help'. A toolbar at the top of the results area includes 'Clear', 'Use WebLicht', 'Export to Personal Workspace', and 'Download'. The search results are displayed in a table with a header 'deu\_news\_2010\_1M, ASV Leipzig'. Each row contains a snippet of text, the location 'Leipzig', and a small circular icon with a question mark.

deu_news_2010_1M, ASV Leipzig		
My Sustainable World wird den in	Leipzig	aufgenommenen, fachübergreifenden Dialog in den Bereichen Energy, Living und Mobility sowie Finanzierung nachhaltiger Konzepte fortführen.
Auch die zweite Mannschaft von Erzgebirge Aue konnte den Siegeszug von RB	Leipzig	nicht stoppen und unterlag zu Hause mit 0:1.
Michael Boris, der erst seit drei Tagen das Zepter als Trainer bei Schalke II schwingt, ging im Vorfeld mit seinen Jungs hart zu Werke: "Gestern in	Leipzig	haben wir kein Spiel gewonnen und nur drei Tore erzielt.
In Richtung	Leipzig	ist die Autobahn allerdings befahrbar.
So wandte sich kürzlich der Lehrer einer neunten Klasse aus	Leipzig	an das NDC.
Der Sprecher von Rasenballsport	Leipzig	, Hans-Georg Felder, sagte, nach der Umgestaltung werde zu erkennen sein, dass es sich um das Stadion von RB Leipzig handle.
Aktuell werden in	Leipzig	parallel zum BMW X1 der 3-Türer, das Coupé und das Cabrio der BMW 1er Rei
Die CDU im Landtag sprach von einem Affront gegen die Messestadt	Leipzig	.

# CLARIN's Virtual Language Observatory



The screenshot displays the CLARIN-D Federated Content Search interface. At the top, the logo features a stylized globe with blue dots and lines, followed by the text "CLARIN-D FEDERATED CONTENT SEARCH". Below this is a search bar containing the text "Leipzig" and a "Search" button. Navigation links for "About", "Search options", "Search results", and "Help" are positioned below the search bar. A toolbar at the top of the results area includes "Clear", "Use WebLicht", "Export to Personal Workspace", and "Download".

The search results are organized into two sections:

- TIGER, IMS, Universität Stuttgart**
  - Text: "Mit der Rheinbraun Verkaufsgesellschaft wurden drei Vertriebstöchter in **Leipzig** , Weimar und Potsdam gegründet , die eine flächendeckende Versorgung mit leichtem Heizöl , Diesel und Schmierstoffen gewährleiten sollen ."
  - Text: "Inzwischen hat sich der Schleier über **Leipzig** und Zwickau , der Lausitz oder dem ' ` schwarzen Dreieck ' ' bei Zittau gelichtet ."
- Dingler Online, Berlin-Brandenburg Academy of Sciences and Humanities**
  - Text: "Vor ungefähr 15 Jahren brachten es die Franzosen unter dem Namen "Chinesische Schminkeblätter," (Rouge en feuille) das erstmalig auf die Messe nach **Leipzig** im Juli 1820. von wo es sich allmählich auch dem östlichen und nördlichen Europa verbreitete."



# Integration in CLARIN

- Integrated in repository of CLARIN center Leipzig
- May function as a template for integration of more CTS servers / instances
- CMDI 1.2 compliant metadata that
  - Allow direct access to services and raw files (here: TEI)
  - Reflects text granularity (currently 3 levels)
    1. Collection (here: Excerpt of *Parallel Bible Corpus*)
    2. Document (here: Bible)
    3. 1 Resource per Book of Bible



# 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

- Planned: FCS endpoint for content search based on existing fulltext index



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	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	...	>
cts	text document	...	>

Mime type: text/xml

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

<< < 1 2 3 4 5 6 > >>

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	...	>

Mime type: text/xml

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

# Digital Athenaeus Index Digger

## Digital Athenaeus

### S. Douglas Olson, Athenaeus' Index of Authors, Texts, and Persons

Where Name='Achaeus of Eretria'

#	Name	Identifier	Sub-Identifier	Prosopographical Reference	Prosopographical Reference id	Work	Edition	Edition id	Passage	Athenaeus Book	Casaubon Reference	Kaibel Reference	Vote	Note Reference	Read Greek Text (Perseus)	Read Greek Text (FrontEnd Uni Leipzig)	Annotate with Perseids
6	Achaeus of Eretria	tragic poet					TrGF	20	F *4	10	10.414c	10.6			<a href="#">10.6</a>	<a href="#">10.6</a>	<a href="#">10.6</a>
7	Achaeus of Eretria	tragic poet					TrGF	20	F *4	10	10.414d	10.6 10.7	<a href="#">Vote</a>		<a href="#">10.6</a> <a href="#">10.7</a>	<a href="#">10.6 - 10.7</a>	<a href="#">10.6 - 10.7</a>
8	Achaeus of Eretria	tragic poet					TrGF	20	F 10	15	15.690b	15.41			<a href="#">15.41</a>	<a href="#">15.41</a>	<a href="#">15.41</a>
9	Achaeus of Eretria	tragic poet					TrGF	20	F 12	4	4.173d	4.74			<a href="#">4.74</a>	<a href="#">4.74</a>	<a href="#">4.74</a>
10	Achaeus of Eretria	tragic poet					TrGF	20	F 13	4	4.173d	4.74			<a href="#">4.74</a>	<a href="#">4.74</a>	<a href="#">4.74</a>
11	Achaeus of Eretria	tragic poet					TrGF	20	F 14	11	11.480f	11.61			<a href="#">11.61</a>	<a href="#">11.61</a>	<a href="#">11.61</a>
12	Achaeus of Eretria	tragic poet					TrGF	20	F 17	14	14.641d	14.48			<a href="#">14.48</a>	<a href="#">14.48</a>	<a href="#">14.48</a>
13	Achaeus of Eretria	tragic poet					TrGF	20	F 19	10	10.461c	10.74			<a href="#">10.74</a>	<a href="#">10.74</a>	<a href="#">10.74</a>
14	Achaeus of Eretria	tragic poet					TrGF	20	F 19	10	10.461d	10.74			<a href="#">10.74</a>	<a href="#">10.74</a>	<a href="#">10.74</a>
15	Achaeus of Eretria	tragic poet					TrGF	20	F 25	6	6.270e	6.99 6.100	<a href="#">Vote</a>		<a href="#">6.99</a> <a href="#">6.100</a>	<a href="#">6.99 - 6.100</a>	<a href="#">6.99 - 6.100</a>
16	Achaeus of Eretria	tragic poet					TrGF	20	F 26	15	15.668a	15.6			<a href="#">15.6</a>	<a href="#">15.6</a>	<a href="#">15.6</a>
17	Achaeus of Eretria	tragic poet					TrGF	20	F 26	15	15.668b	15.6			<a href="#">15.6</a>	<a href="#">15.6</a>	<a href="#">15.6</a>
18	Achaeus of Eretria	tragic poet					TrGF	20	F 27	7	7.277b	7.5			<a href="#">7.5</a>	<a href="#">7.5</a>	<a href="#">7.5</a>

[http://www.digitalatheneus.org/tools/OlsonIndex/digger.php?what\[\]=Name|Achaeus+of+Eretria&onoffswitch=on](http://www.digitalatheneus.org/tools/OlsonIndex/digger.php?what[]=Name|Achaeus+of+Eretria&onoffswitch=on)

# Interoperability through application independent references

Re-engineering the brain

הסתכלו אף צילית של שות עשויה ליראמא...  
בירמיהו, ול עקרה עורמה...  
היא אר-עבב אדא...  
הקבלה מתולה לענ...  
בכל פעם עפמ עפמא את המשל...  
עמם פאן 10,000 ת-הי-עב-ב...  
קנ שוטמ ויא ק קי קילק...  
שאר שרמ קסמא פול קילק...  
המחמת פולמ 100 פ"ק עמ...  
יהוד תורבוקס...  
התקבוקס פולמ...  
הירשוט פולמ עורעד...  
התנבוקס פולמ עורעד...  
אבל, דרועס קרין...  
בסוג עפמ אל דרועס כידר לערוח את התנבוקס...  
אינם בסוגים את יהוד שמחמת עפמחמו ב...  
פרי, קלעקוקס...  
על ק לעמא את יהוד...  
לעל כידר...  
מחמת תדוקס מחמת אפי' ויה...  
פרי, לילתת תו פרוסו הספיקס כידר...  
פחן, עפמית יתי כילתת ערמון איהם...  
קאמו אתם פרוסו תרמכום...  
זל קל תודר, לילתת

Copyright 2014 Leipzig University / Martin Reichle

Digital Athenaeus

S. Douglas Olson, Athenaeus' Index of Authors, Texts, and Persons

Where Name=Achaeus of Eretria

	#	Identifier	Sub-Identifier	Author	Text	Page	Classification	Creation Reference	Final Reference	Initial Reference	Text Date (Person)	Text Date (Language)	Analysis with Person
	100	000		Achaeus of Eretria	De Generationibus	100		10.0	10.0		10.0	10.0	
	101	000		Achaeus of Eretria	De Generationibus	101		10.1	10.1		10.1	10.1	

Virtual Language Observatory

Record: The Bible in French, King James Version.

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

Technically, these unrelated projects can now work with shared resources.

# Why include support for CTS URNs?

CTS is developed by Humanists & reflects the requirements from certain Digital Humanity communities

Perseus, Croatiae Auctores Latini, CITE, ...

Creating CTS ready data is a research & project goal in DH

“pick researchers up, where they are“

Technical benefits

Normalisation of generic text access, “Outsourcing“ of text content, The Canonical Text Infrastructure, ...

# Contact

Jochen Tiepmar

E-Mail: [jtiepmar@informatik.uni-leipzig.de](mailto:jtiepmar@informatik.uni-leipzig.de)

Scalable Data Solutions (ScaDS) Leipzig

Universität Leipzig

Ritterstraße 9-13

04109 Leipzig

