

The Billion Words Library

Canonical Text Services (CTS)

Jochen Tiepmar

NLP-Group

University of Leipzig

CTS

- Protocol specifying a online text service for citation
- developed in the homermultitext project (www.homermultitext.org), Smith et.al.2009
<http://www.homermultitext.org/hmt-docs/specifications/ctsum/>
<http://www.homermultitext.org/hmt-docs/specifications/cts/>
- Unique Resource Name (URN) identifies text passage

Testlab: www.urncts.de

Example:

About Christmas Trees

According to E. Anschütz the leaves of a christmas tree are verdant in the winter (1).

(1) <urn:cts:songs:christmas.ohtennenbaum.en.1:1.1.1-1.1.4>

URN / Basic Idea

urn:cts:[CTS_NAMESPACE]:[WORK]:[PASSAGE]

[WORK] = TEXTGROUP.WORK.VERSION.EXEMPLAR

document from the outside, **outer structure**

Shakespeare → **Sonnets** → **english** → **1st edition**

Shakespeare → **Hamlet** → **german** → **2nd edition**

[PASSAGE]

passage in the document, **inner structure**

Sonnet 35

Sonnet 35 → **Line 1**

Sonnet 35 → **Line 1 to Sonnet 35** → **Line 5**

Shakespeare → **Sonnets** → **english** → **1st edition** → **Sonnet 35** → **Line 1**

URN / Examples

urn:cts:demo:shakespeare.sonnets:

Shakespeares Sonnets

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

german translation

urn:cts:demo:shakespeare.sonnets:35.1

line 1 in sonnet 35

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

line 1 to 5 in sonnet 35

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

line 1 word „grieved“ to line 5 first occurrence of „faults“

Functions

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"/>
```

Functions

GetCapabilities ()

GetValidReff (urn, level)

All URNs belonging to [urn] with distance [level] → all URNs from chapter 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>
```

Functions

GetCapabilities ()

GetValidReff (urn, level)

GetFirstUrn (urn)

First URN that belongs to [urn] → 1st URN in chapter 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>
```

Functions

GetCapabilities ()

GetValidReff (urn, level)

GetFirstUrn (urn)

GetPrevNextUrn (urn)

Left and right URNs from [urn] → URNs left and right of line 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>
```


Functions

GetCapabilities ()

GetValidReff (urn, level)

GetFirstUrn (urn)

GetPrevNextUrn (urn)

GetLabel (urn)

Informal description of [urn] → „Shakespears Sonnet 36 Line 9“

GetPassage (urn)

GetPassagePlus (urn)

```
<cts:GetLabel>  
- <cts:request>  
  <cts:requestName>GetLabel</cts:requestName>  
  <cts:urn>urn:cts:latinlit:phi1014.phi001.lat1:1.1</cts:urn>  
</cts:request>  
- <cts:reply>  
  <cts:label>Edition:Controversiae</cts:label>  
</cts:reply>  
</cts:GetLabel>
```

Functions

GetCapabilities ()

GetValidReff (urn, level)

GetFirstUrn (urn)

GetPrevNextUrn (urn)

GetLabel (urn)

GetPassage (urn)

Passage specified by [urn] → the 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 occidit;
    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 lupan
    FVLVI Sparsi.
  </cts:passage>
</cts:reply>
</cts:GetPassage>
```

Functions

GetCapabilities ()

GetValidReff (urn, level)

GetFirstUrn (urn)

GetPrevNextUrn (urn)

GetLabel (urn)

GetPassage (urn)

GetPassagePlus (urn)

All of the above except the textinventory

```
<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 sa
  occidit; at hercule lenonem non occidisti. Deducta es in lupanar, ac
  obscenum lectulum uocas ? de pudicitia sacerdotis hic quaeritur. "T
  mihi sacerdotem cuius precaria est castitas? Cum ex illo lupanari c
  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>
```

Http-Requests / [Server]

GetCapabilities

`http://urncts.de/demo/cts/cts5`

GetPassage:

`http://urncts.de/demo/cts/cts5`

GetValidReff

`http://urncts.de/demo/cts/cts5`

Http-Requests / [Server]?[request]

GetCapabilities

<http://urncts.de/demo/cts/cts5?request=GetCapabilities>

GetPassage:

<http://urncts.de/demo/cts/cts5?request=GetPassage>

GetValidReff

<http://urncts.de/demo/cts/cts5?request=GetValidReff>

Http-Requests / [server]?[request]&[parameters]

GetCapabilities

<http://urncts.de/demo/cts/cts5?request=GetCapabilities>

GetPassage:

<http://urncts.de/demo/cts/cts5?request=GetPassage&urn=urn:cts:demo:goethe.faust2.de:1.2>

GetValidReff

<http://urncts.de/demo/cts/cts5?request=GetValidReff&urn=urn:cts:demo:goethe.faust2.de:&level=2>

(...)

(...)

Multilanguage

UTF-8

```
<div1 n="1" type="multilang">
  <div2 n="1" xml:lang="de" type="singlelang"> Das Haus auf dem Berg. äöüÄÖÜß. </div2>
  <div2 n="2" xml:lang="ar" type="singlelang"> داس هاوس عوف ماركا بيرغ </div2>
  <div2 n="3" xml:lang="zh(tradition)" type="singlelang"> 達斯豪斯奧夫DEM伯格 </div2>
  <div2 n="4" xml:lang="zh(simple)" type="singlelang"> 达斯豪斯奧夫DEM伯格 </div2>
  <div2 n="5" xml:lang="gu" type="singlelang"> દાસ હાઉસ ઓફ ડેમને બર્ગ </div2>
  <div2 n="6" xml:lang="gr" type="singlelang"> Το σπίτι στο λόφο </div2>
  <div2 n="7" xml:lang="am" type="singlelang"> ስጢን ሞሃር ህንጻ </div2>
  <div2 n="8" xml:lang="bn" type="singlelang"> পাহাড় ঘর </div2>
  <div2 n="9" xml:lang="bg" type="singlelang"> Къщата на хълма </div2>
  <div2 n="10" xml:lang="he" type="singlelang"> הבית על הגבעה </div2>
  <div2 n="11" xml:lang="in" type="singlelang"> पहाड़ी पर घर </div2>
  <div2 n="12" xml:lang="jp" type="singlelang"> 丘の上の家 </div2>
  <div2 n="13" xml:lang="ji" type="singlelang"> די הויז אויף די בערגל </div2>
  <div2 n="14" xml:lang="Khmer" type="singlelang"> ផ្ទះនៅលើភ្នំនេះ </div2>
  <div2 n="15" xml:lang="kr" type="singlelang"> 언덕에 집 </div2>
  <div2 n="16" xml:lang="lo" type="singlelang"> ຄື້ອນງົງວັບໜ່ວຍຸ </div2>
  <div2 n="17" xml:lang="mr" type="singlelang"> टेकडी वर घर </div2>
  <div2 n="18" xml:lang="ne" type="singlelang"> पहाडी मा घर </div2>
  <div2 n="19" xml:lang="Persisch" type="singlelang"> خانه در تپه </div2>
  <div2 n="20" xml:lang="Punjabi" type="singlelang"> ਪਹਾੜੀ 'ਤੇ ਘਰ ਠੁੰ </div2>
  <div2 n="21" xml:lang="ru" type="singlelang"> Дом на холме </div2>
  <div2 n="22" xml:lang="tl" type="singlelang"> Լժանը Լիճք վրա </div2>
  <div2 n="23" xml:lang="te" type="singlelang"> కొండ మీద వూస్ </div2>
  <div2 n="24" xml:lang="th" type="singlelang"> บ้านอยู่บนเนินเขา </div2>
  <div2 n="25" xml:lang="ur" type="singlelang"> پہاڑی پر گھر </div2>
  <div2 n="26" xml:lang="vn" type="singlelang"> Ngôi nhà trên đồi </div2>
</div1>
```


Additional Usecases

1 Using URNs as IDs

2 Standardisation

3 Structurebased translation

4 Meta information

5 Change text dynamically

2-5 limited to sql based implementation

2-5 not specified/wanted by CTS, but also not forbidden

3 depends on data

Additional Usecases / URNs as IDs

(similar to CITE-Architecture <http://www.homermultitext.org/hmt-docs/cite/>)

URNs provide unique identifiers down to character/sign-level

Mapping between URN and Integer-ID, picture, resource

Additional Usecases / Standardization

```
<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>
    ...
  </div2>
</div1>
</passage>
```

The screenshot displays the CTS Browser interface. At the top, a search bar contains the text "demo - songs". Below the search bar, a list of search results is shown, including "christmas", "ohtennenbaum", and "cat - de - en - fr - it - lat". The main content area shows the rendered output of the TEI XML document, which is a list of lines of text grouped into three strophes (1, 2, and 3). The first strophe contains six lines, the second contains six lines, and the third contains six lines. The rendered text is as follows:

1	1	O Tannenbaum, O Tannenbaum,
	2	Wie treu sind deine Blätter.
	3	Du grünst nicht nur zur Sommerzeit,
	4	Nein auch im Winter wenn es schneit.
	5	O Tannenbaum, O Tannenbaum,
	6	Wie grün sind deine Blätter!
1	1	O Tannenbaum, O Tannenbaum,
	2	Du kannst mir sehr gefallen!
	3	Wie oft hat schon zur Winterszeit
	4	Ein Baum von dir mich hoch erfreut!
	5	O Tannenbaum, O Tannenbaum,
	6	Du kannst mir sehr gefallen!
1	1	O Tannenbaum, O Tannenbaum,
	2	Dein Kleid will mich was lehren:
	3	Die Hoffnung und Beständigkeit
	4	Gibt Mut und Kraft zu jeder Zeit!
	5	O Tannenbaum, O Tannenbaum,
	6	Dein Kleid will mich was lehren.

At the bottom right of the rendered output, there is a logo for "BSV" (Bibliographic Service) and the text "CTS Browser by Martin Reckziegel Image Processing/BSV University of Leipzig".

Additional Usecases / structurebased translation

`<cts:urn>urn:cts:songs:christmas.ohtennenbaum.en.1:1.1.2</cts:urn>`

`<cts:passage>How are thy leaves so verdant !</cts:passage>`

`<cts:urn>urn:cts:songs:christmas.ohtennenbaum.fr.1:1.1.2</cts:urn>`

`<cts:passage>Que j'aime ta verdure !</cts:passage>`

`<cts:urn>urn:cts:songs:christmas.ohtennenbaum.lat.1:1.1.2</cts:urn>`

`<cts:passage>Fidissima cum fronde,</cts:passage>`

`<cts:urn>urn:cts:songs:christmas.ohtennenbaum.de.1:1.1.2</cts:urn>`

`<cts:passage>Wie treu sind deine Blätter.</cts:passage>`

`<cts:urn>urn:cts:songs:christmas.ohtennenbaum.it.1:1.1.2</cts:urn>`

`<cts:passage>risplendi nella notte !</cts:passage>`

Additional Usecases / structurebased translation

```
<cts:urn>urn:cts:dta:09.343012.de.transcript:2</cts:urn>
```

```
<cts:passage>
```

Jahres/ vnd den 31. dieses in die Closter Kirche mit Christlichen Ceremonien bestattet worden; Gethan I

```
</cts:passage>
```

```
<cts:urn>urn:cts:dta:09.343012.de.norm:2</cts:urn>
```

```
<cts:passage>
```

Jahres/ und den 31. dieses in die Kloster Kirche mit Christlichen Zeremonien bestattet worden; Getan I

```
</cts:passage>
```

```
<cts:urn>urn:cts:dta:09.343012.de.translit:2</cts:urn>
```

```
<cts:passage>
```

Jahres/ vnd den 31. dieses in die Closter Kirche mit Christlichen Ceremonien bestattet worden; Gethan D

```
</cts:passage>
```

Additional Usecases / meta information

```
<passage>  
- <div1 n="1" type="song">  
  - <div2 n="1" type="strophe">  
    <div3 n="1" letters="24" tokens="4" avg_tokensize="6" type="line"> O Tannenbaum, O  
    <div3 n="2" letters="24" tokens="5" avg_tokensize="4" type="line"> Wie treu sind deine  
  </div2>  
</div1>  
</passage>
```

Additional Usecases / dynamic text changes

Normalising texts, stemming, delete special characters

`<passage>`

...

`<div3 n="2" type="line">` How are thy leaves so verdant ! `</div3>`

`<div3 n="2" type="line">` HOW ARE THY LEAVES SO VERDAND

`<div3 n="2" type="line">` ! dnadrev os sevaeI yht era woH `</div3>`

– `<div3 n="2" type="line">`

H.o.w.a.r.e.t.h.y.l.e.a.v.e.s.s.o.v.e.r.d.a.n.t.!

`</div3>`

– `<div3 n="2" type="line">`

.....|---|..--||.-|.-|..||-|.....|---||.-...||.-|...-|.....||...|---||...-|..-|.-|.-|.-|.-|..||!

`</div3>`

...

`</passage>`

```
<cts:GetPassage>
```

```
– <cts:request>
```

```
  <cts:requestName>GetPassage</cts:requestName>
```

```
  <cts:urn>urn:cts:presentation:last.slide.en:</cts:urn>
```

```
</cts:request>
```

```
– <cts:reply>
```

```
  – <cts:passage>
```

```
    <div1 n="1" type="line">Thank you for your attention.</div1>
```

```
    <div1 n="2" type="line">Do you have any questions?</div1>
```

```
  </cts:passage>
```

```
</cts:reply>
```

```
</cts:GetPassage>
```