

# Canonical Text Service

Jochen Tiepmar

ScaDS, ASV, Uni Leipzig

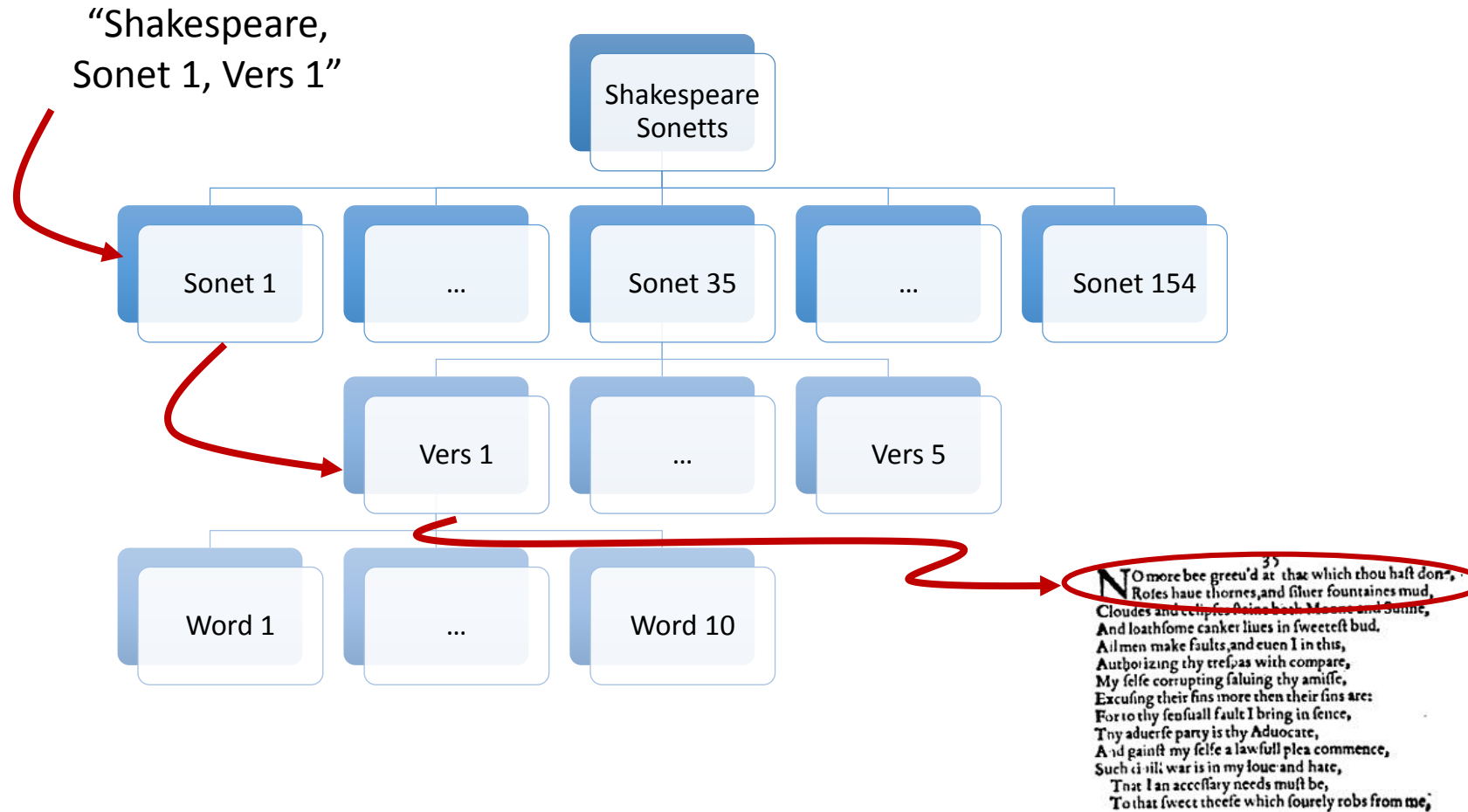


# Overview

## Canonical Text Services (CTS)

- protocol for a webbased citable text service
- Unique Identifiers(**U**nique **R**esource **N**ame, **URN**) refer to text passages
- Developed in Homer Multitext Project([www.homermultitext.org](http://www.homermultitext.org)), Smith et.al.2009  
<http://www.homermultitext.org/hmt-docs/specifications/ctsur/>  
<http://www.homermultitext.org/hmt-docs/specifications/cts/>
- This implementation was done in Billion Words Project
- Implementation for Tripelstore and XML-DB not suitable for BW-Project
- Demo webpage: [www.urncts.de](http://www.urncts.de)

# Hierarchical Documents



# Citation

## „Outer“ Hierarchy

Shakespeare → Sonnets → english → 1st edition

## „Inner“ Hierarchy

Sonnet 1 → Vers 1

## Combination

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

## CTS-URN

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

# Canonical Text Services (CTS)

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

“From fairest  
creatures we desire  
increase,”

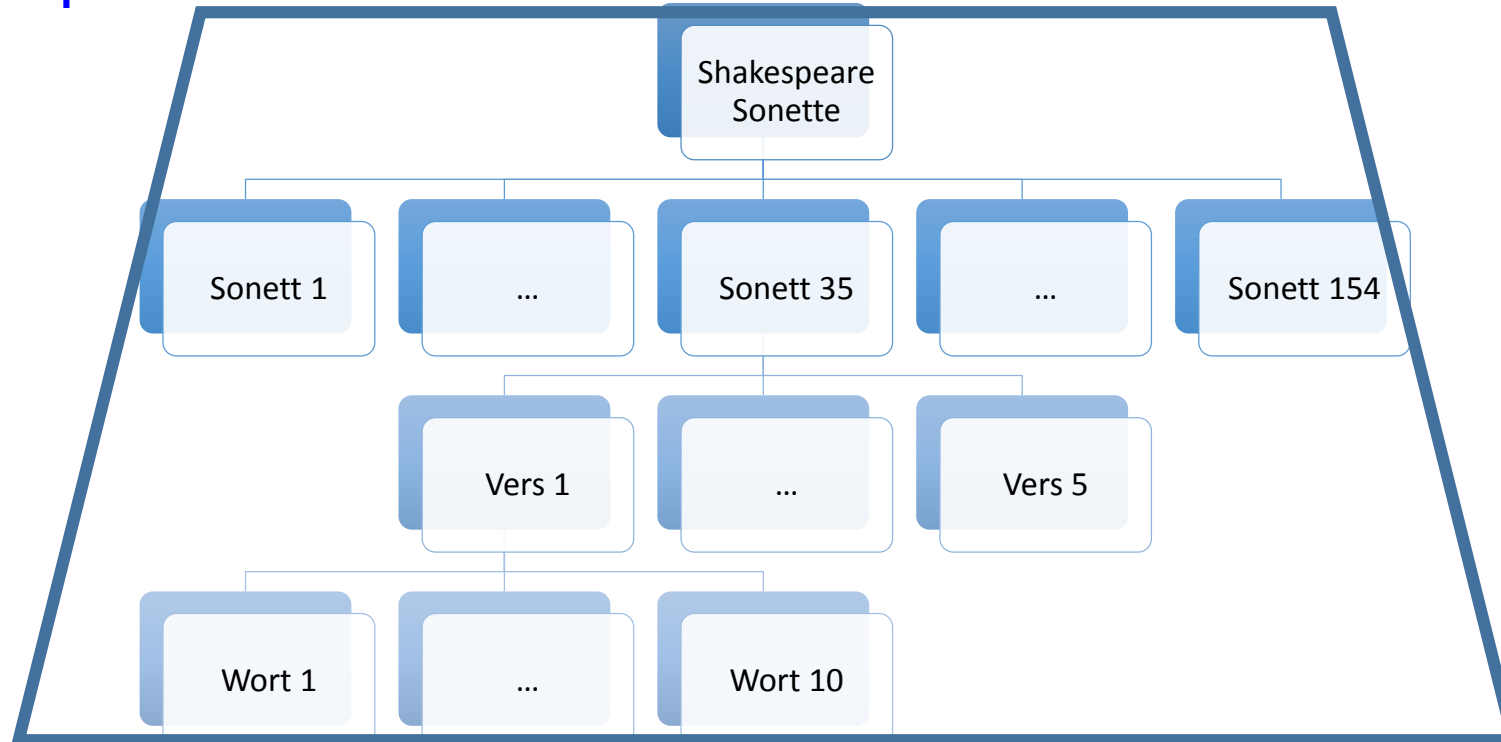


CTS

# Kanonische Zitation

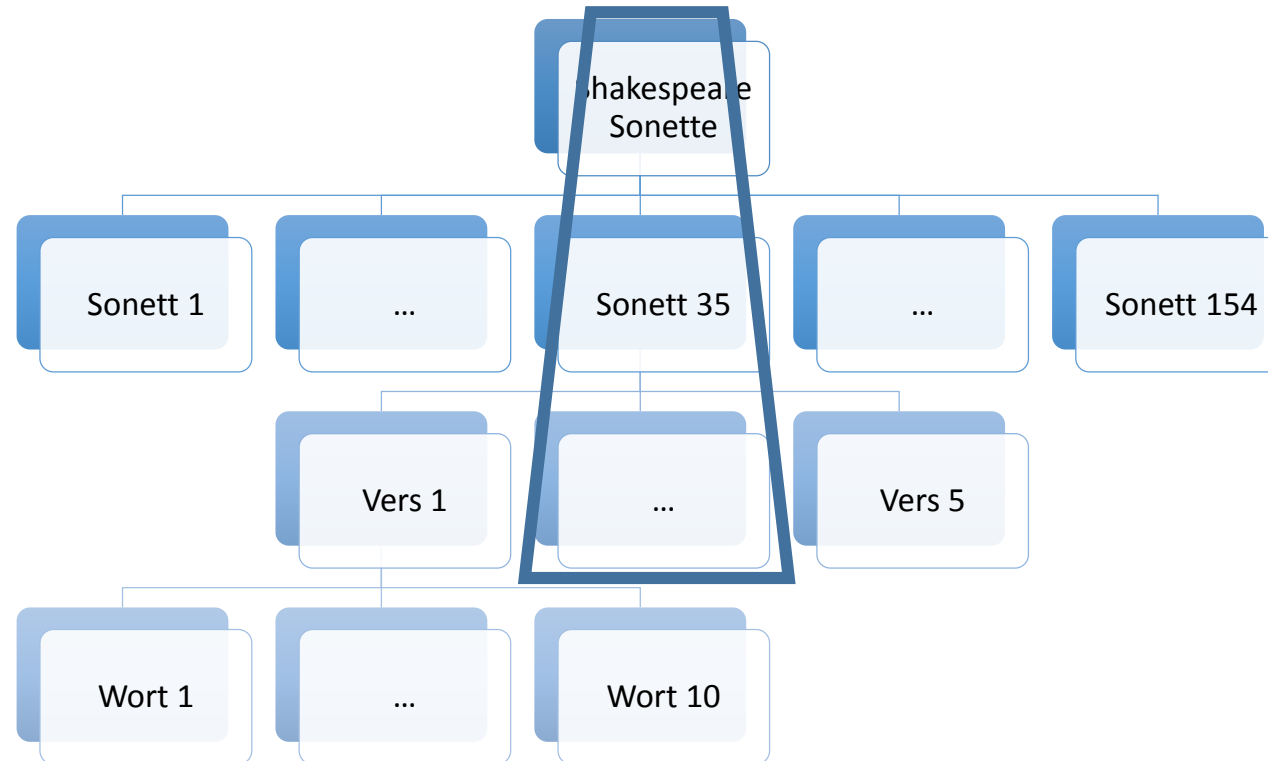
urn:cts:demo:[shakespeare.sonnets:](#)

urn:cts:demo:[shakespeare.sonnets.de:](#)



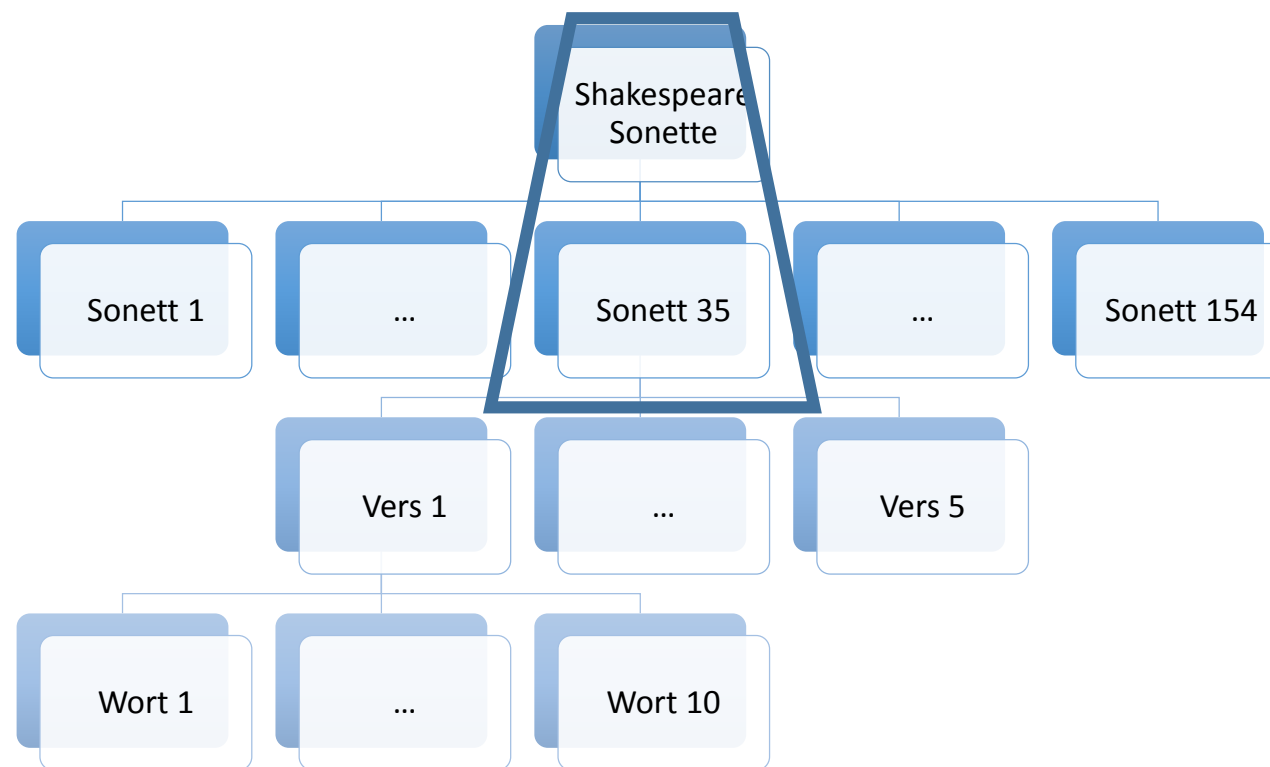
# Kanonische Zitation

urn:cts:demo:shakespeare.sonnets:35.4



# Kanonische Zitation

urn:cts:demo:shakespeare.sonnets:35

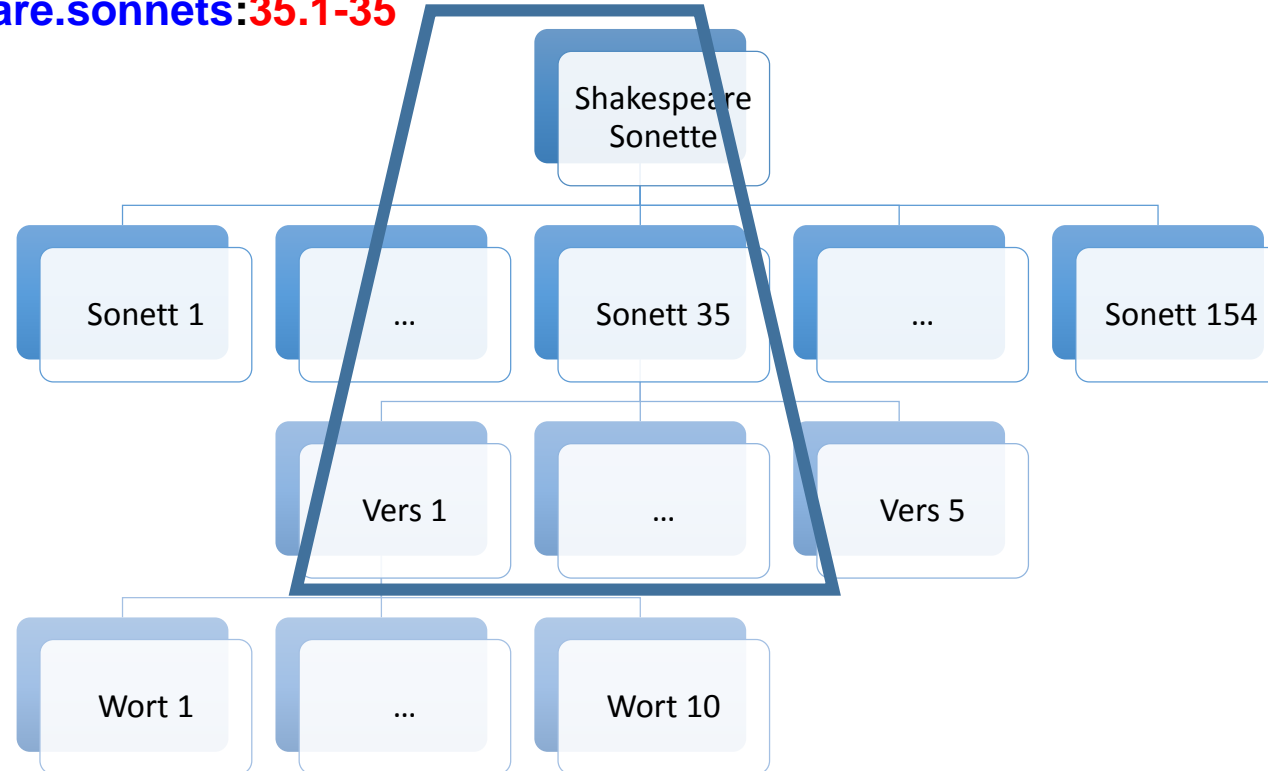




# Kanonische Zitation

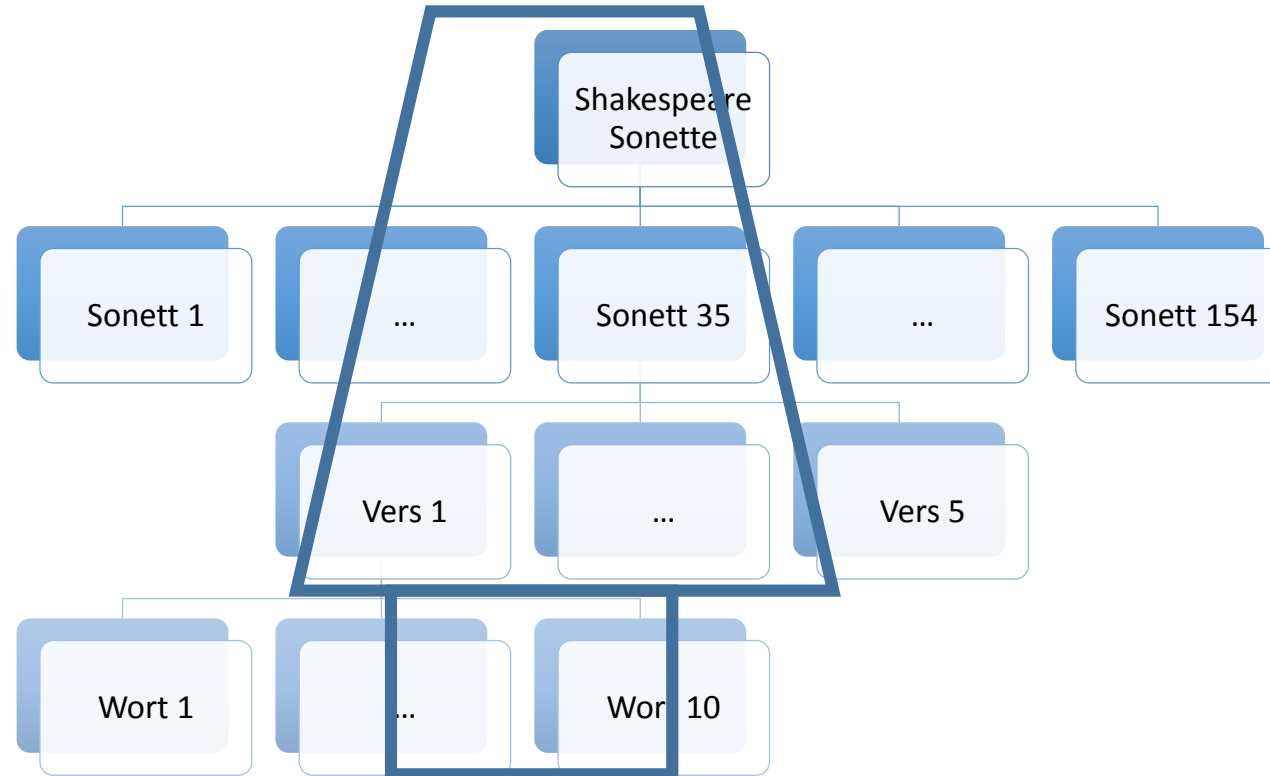
urn:cts:demo:shakespeare.sonnets:35.1-35.5

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



# Kanonische Zitation

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



# Functions

**GetCapabilities ( )**

**GetValidReff ( urn, level )**

**GetFirstUrn ( urn )**

**GetPrevNextUrn ( urn )**

**GetLabel ( urn )**

**GetPassage ( urn )**

**GetPassagePlus ( urn )**

**Function name & Parameters as GET in HTTP-Request**

# 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] → 1<sup>st</sup> 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 „neighbour“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“

```
Get <GetLabel>  
  <request>  
    <requestName>GetLabel</requestName>  
    <requestUrn>urn:cts:pbz: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>
```



# 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 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>
```

# 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 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 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>
```

# CTS-Iterator

Mapping HTTP Requests and XML Documents on Iterator functions

-> More attractive for Developers

JAVA and Ruby-Wrapper done, more „when it's done“

[Open Source](#)

```
CTSAccess cts = new CTSAccess();
String url = "http://ctstest.informatik.uni-leipzig.de/dta/cts/";
cts.setBaseUrl(url);
EditionIterator it = cts.getEditionIterator();
while(it.hasNext()){
    Edition e = it.next();
    UrnIterator urns = e.getUrnIterator();
    while(urns.hasNext()) || while(urns.hasNext(2))
    {
        String urn = urns.next(); || String urn = urns.next(2);
        System.out.print("->" + cts.getLabel(urn) + "->" + cts.getPassage(urn));
    }
}
```

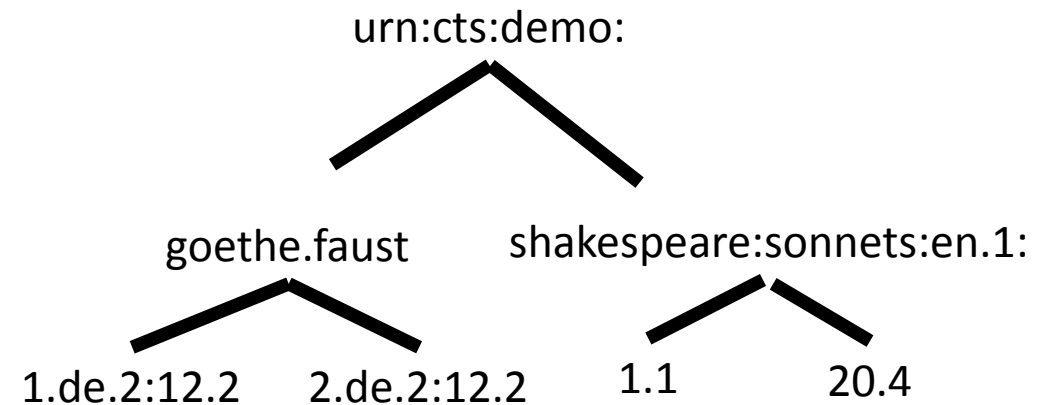
# Data structure

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

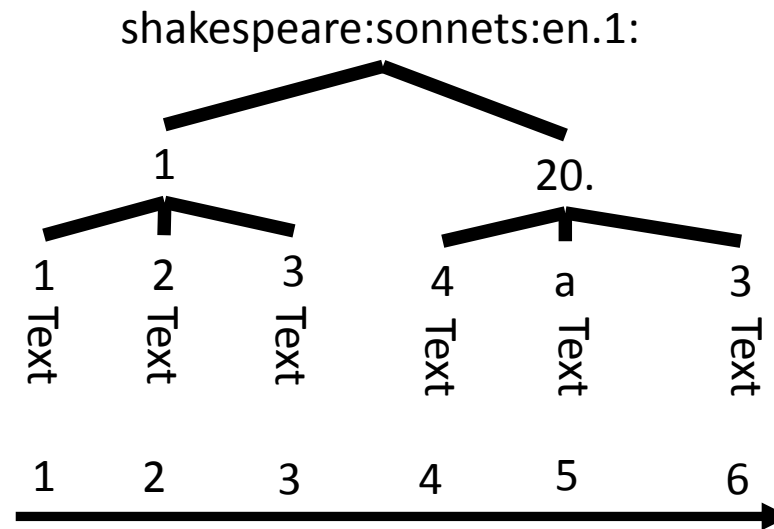
**urn:cts:demo:shakespeare.sonnets.en.1:20.4**

**urn:cts:demo:goethe.faust1.de.2:12.2**

**urn:cts:demo:goethe.faust2.de.2:12.2**



# Data structure

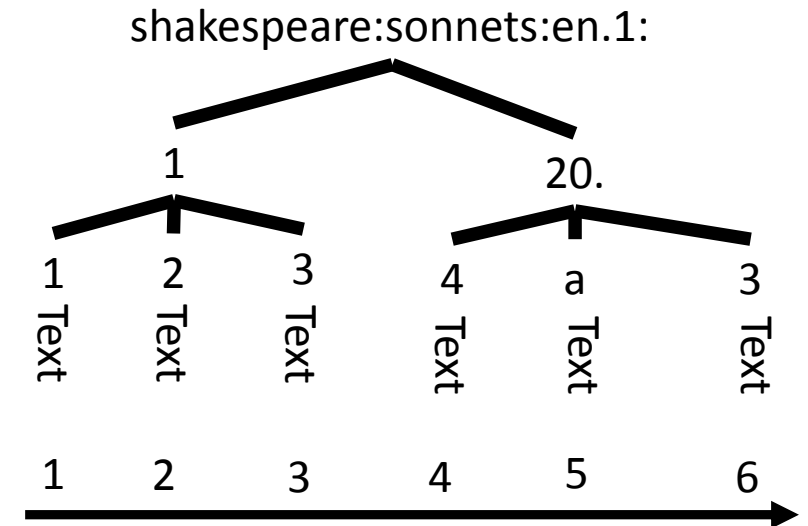


# Data structure - Advantages

## Prefix Optimisation

urn:cts:demo:shakespeare.sonnets.en.1:1.1  
\_20.4  
\_goethe.faust1.de.2:12.2  
\_2.de.2:12.2

## Logarithmical search times



# Data

	Tokens	
DTA, Deutsches Text Archive	334'820'482	Various german texts each in 3 editions
PBC, Parallel Bible Corpus	247'292'629	831 Bible translations
Perseus	27'295'030	greekLit, latinLit, farsiLit, pdlrefwk
100k	1'281'272'600	Randomly generated documents

# DTA

- BBAW
- 5'136 Editions
- 3 Editions per Document (translit, transcript, norm)
- 1 Citationlevel (Sentence)

[Kafka](#), [Goethe](#), [Kant](#), [Gauss](#)



# PBC

- 831 Editions
- Bible translations
  - [5 german bibles](#)
  - 3 Citationlevels (Book, Chapter, Sentence)

# Perseus

- Alphios CTS
- 1'175 Editions
- [greekLit](#), [latinLit](#), [farsiLit](#), [pdlrefwk](#)
- Heterogenely structured

# Statistics

## Test PC:

- Ubuntu-Server
- VM in Universities network

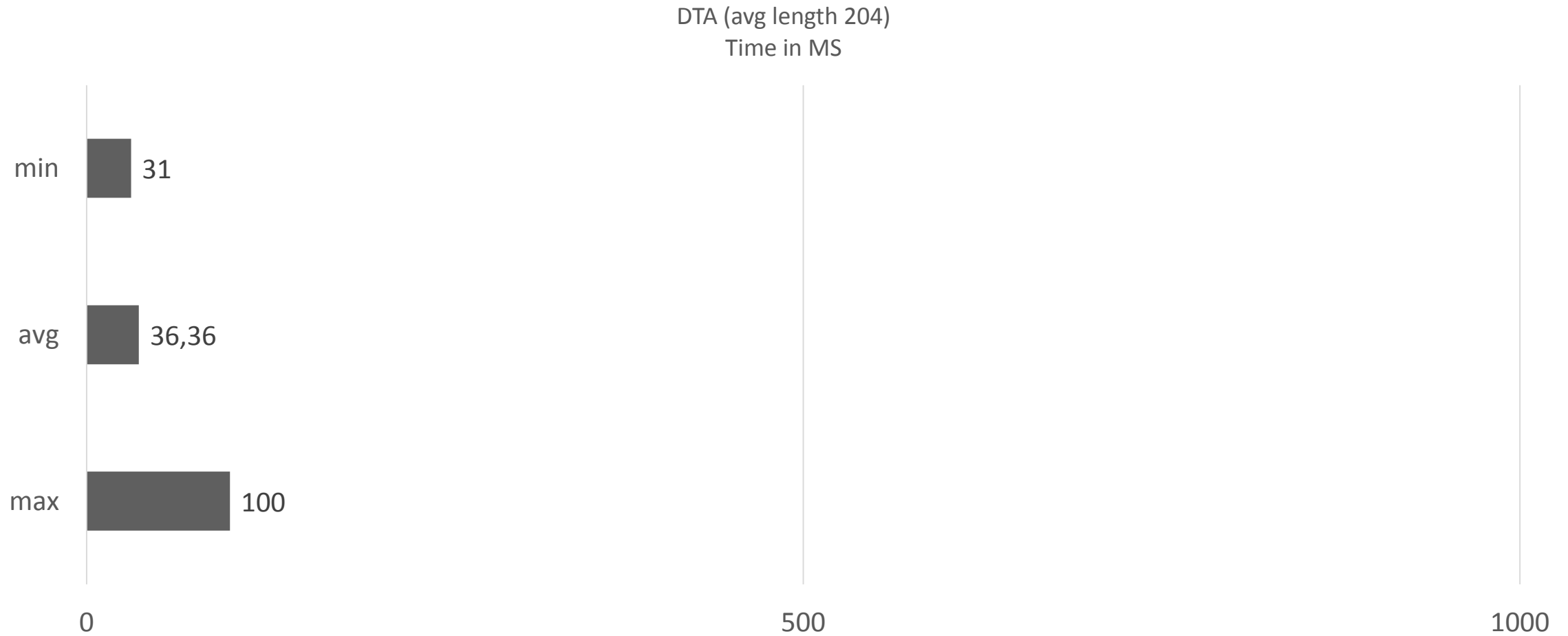
## Testsetup:

- Get all editions
- Request the passage spanning the first 2 URNs

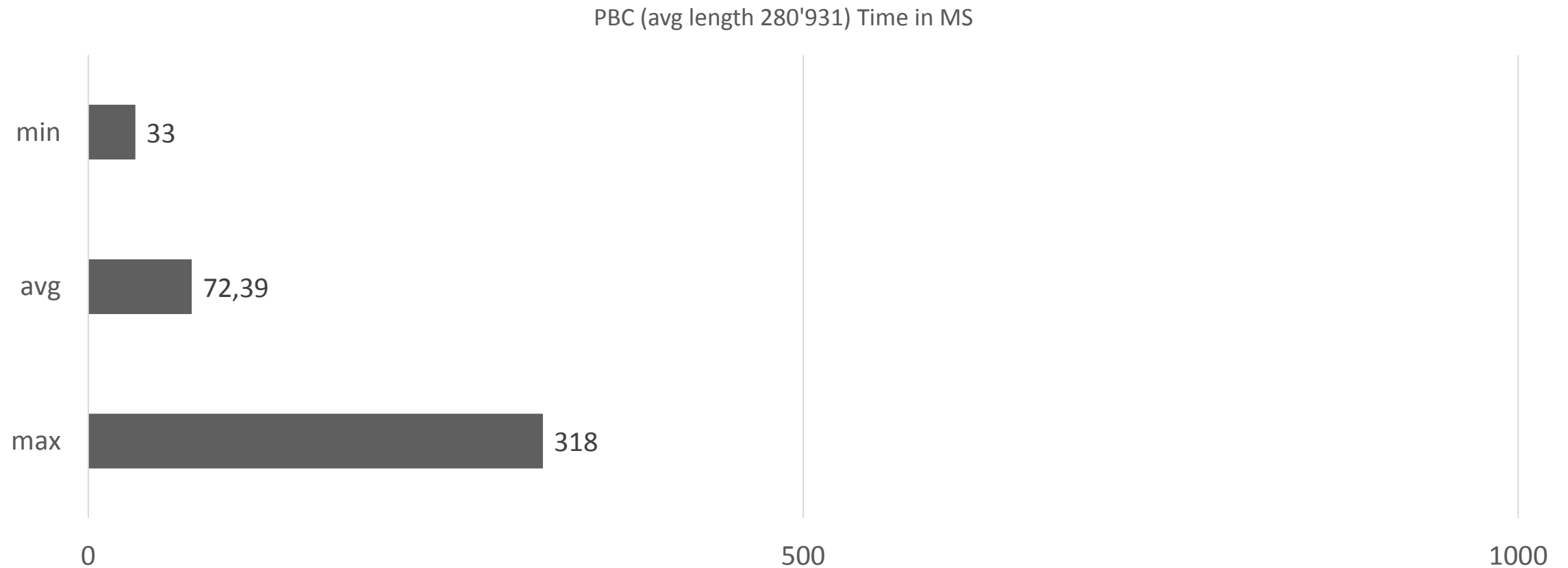
```
ctsuser@ctstestbed:~$ free -h
              total        used        free      shared    buffers     cached
Mem:           1.0G          794M          207M           0B          109M          252M
-/+ buffers/cache:          432M          570M
Swap:           509M           21M          488M

ctsuser@ctstestbed:~$ cat /proc/cpuinfo
processor       : 0
vendor_id      : AuthenticAMD
cpu family     : 15
model          : 6
model name     : Common KVM processor
stepping       : 1
microcode      : 0x1000065
cpu MHz        : 2399.998
cache size     : 512 KB
fdiv_bug       : no
hlt_bug        : no
f00f_bug       : no
coma_bug       : no
fpu            : yes
fpu_exception  : yes
cpuid level    : 5
wp             : yes
flags          : fpu de pse tsc msr pae mce cx8 api
bogomips       : 4799.99
clflush size   : 64
cache_alignment : 64
address sizes   : 40 bits physical, 48 bits virtual
```

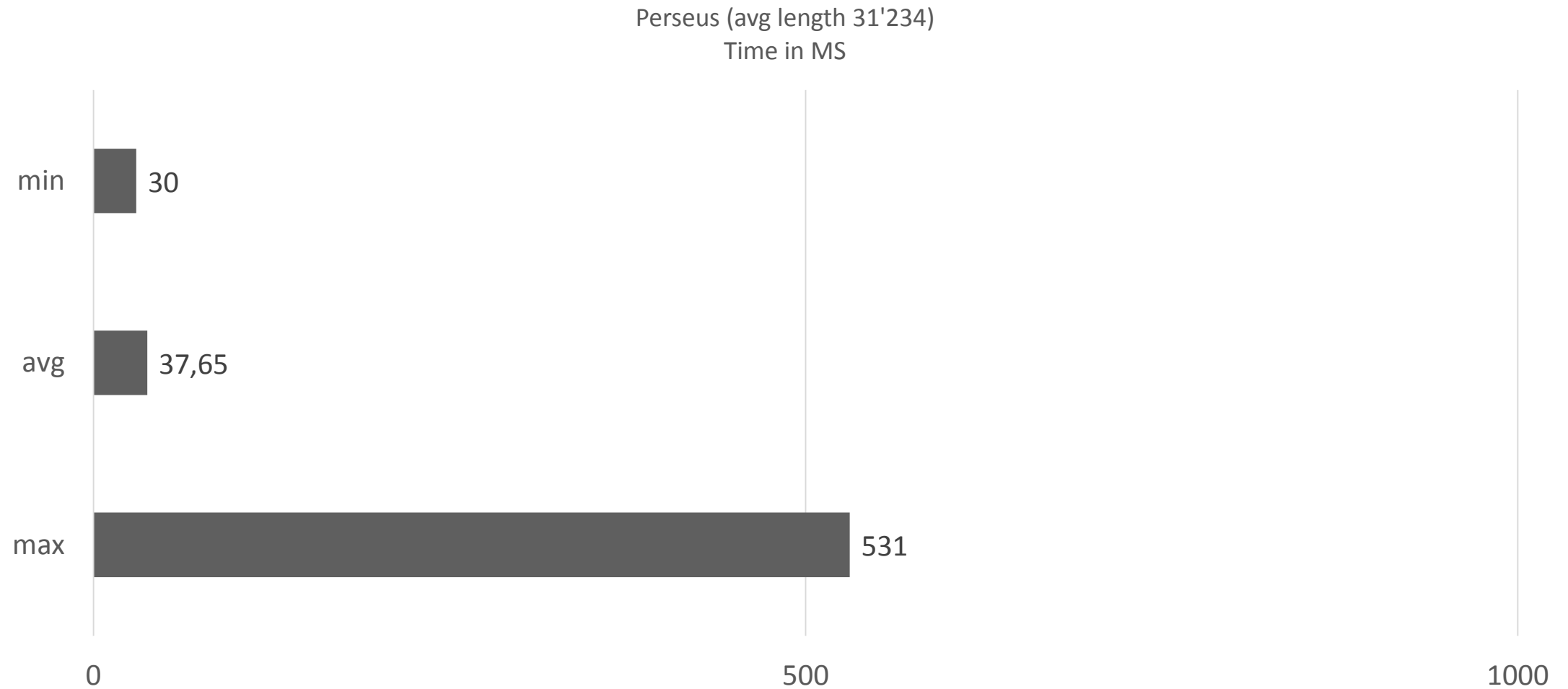
# DTA Benchmarks



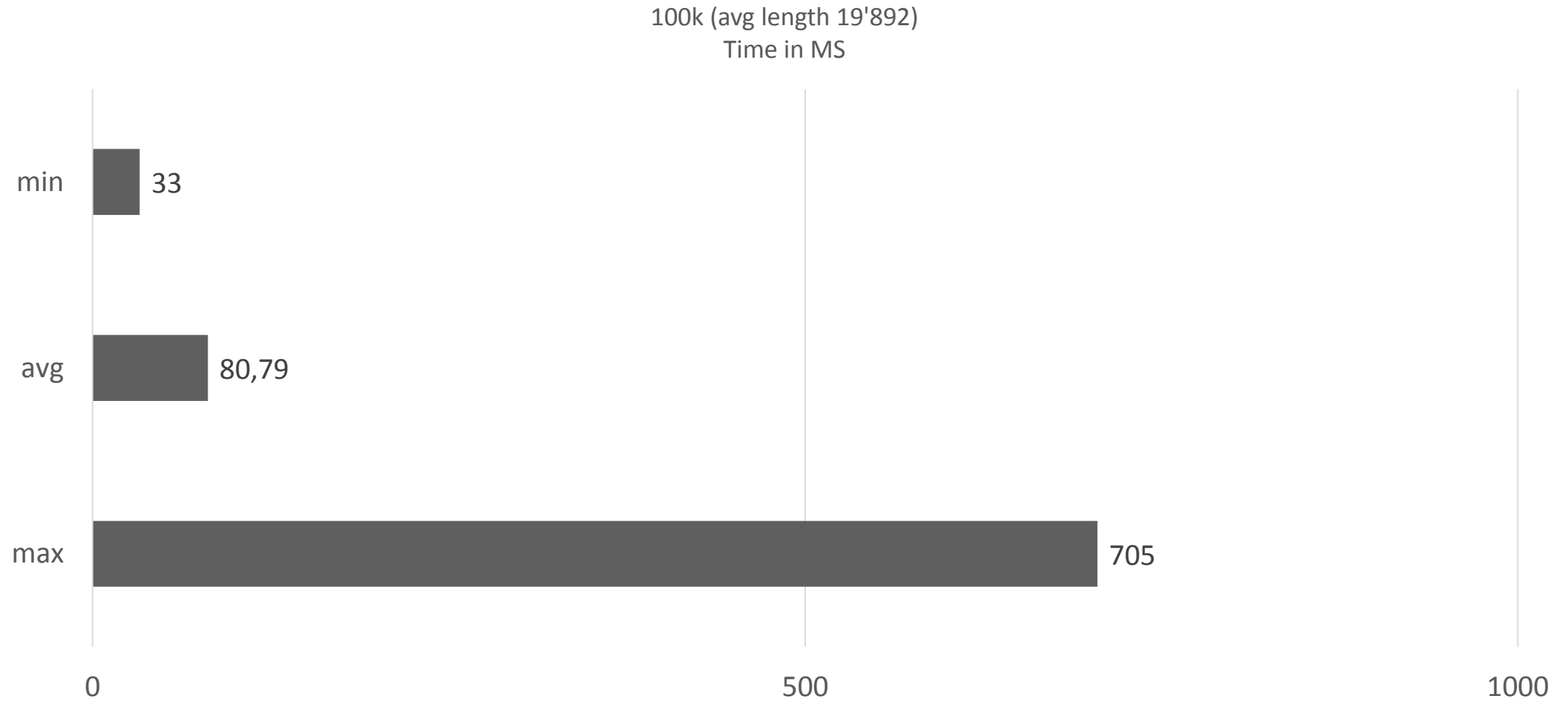
# PBC Benchmarks



# Perseus Benchmarks



# 100k Benchmarks



# Statistics

	Shortest 1/4 (MS)	Longest 1/4 (MS)
DTA	36,00	37,29
PBC	60,70	91,50
Perseus	33,76	47,86
100K	78,64	83,08

	Shortest 1/10 (MS)	Longest 1/10 (MS)
DTA	35.90	37,93
PBC	56,62	98,05
Perseus	33,59	60,24
100K	75,31	81,51



# Advantages of CTS

## **Online Access**

- Central Dataset**

- Less Redundancy**

- Mapping large texts on small URNs**

## **Standardisation**

- Requests independent from name of text units (Chapter, Song, Book)**

- Text passage always in<passage> - element**

# Advantages of this CTS

## Online Access

- Central Dataset

- Less Redundancy

- Mapping large texts on small URNs

## Standardisation

- Requests independent from name of text units (Chapter, Song, Book)

- Text passage always in<passage> - element

## Configurable Views

- Passage is constructed from its text units -> „rules for construction“ can be changed to present diff. Views**

- Views per GET-Parameter at runtime**

- Standardisation of text units creates access points for generic tools**

# Views - Plain Text

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>

# Views - <div>

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

```
<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>
```

# Views - EpiDoc

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

```
<passage><tei:TEI><tei:text><tei:body>
  <tei:div n="1" type="song">
    <tei:div n="1" type="strophe">
      <l n="1">O Tannenbaum, O Tannenbaum, </l>
      (...)
      <l n="6">Wie grün sind deine Blätter! </l>
    </tei:div>
    <tei:div n="2" type="strophe">
      <l n="1">O Tannenbaum, O Tannenbaum, </l>
      (...)
      <l n="4">Ein Baum von dir mich hoch erfreut!</l>
    </tei:div>
  </tei:div>
</tei:body></tei:text></tei:TEI></passage>
```

# Views - Metainformation

urn:cts:songs:christmas.ohtennenbaum.de.1:1.1.1

<passage>

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

<passage>

VS.

<passage>

<div3 n="1" type="line" letters="24" tokens="4" avg\_tokensize="6">O Tannenbaum, O  
Tannenbaum, </div3>

<passage>

# Views - Various

## **DeleteXML**

**Strip XML from text**

## **EscapePassage**

**Escape the passage**

## **FormatXML**

**Format XML in passage**

# Use Cases

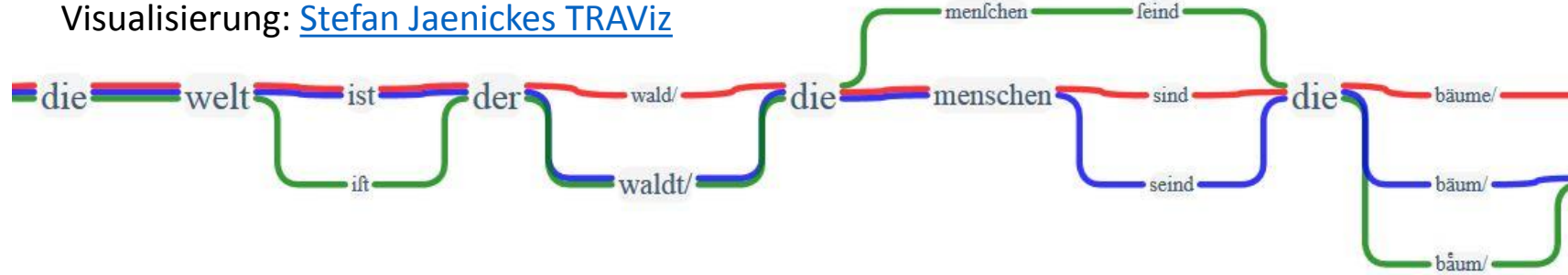
- > **Alignment**
- > **Generic tools and reader**
- > **CTS – Text Miner**



# Alignment

urn:cts:dta:albertinus.landtstoertzer011615.de.translit:11

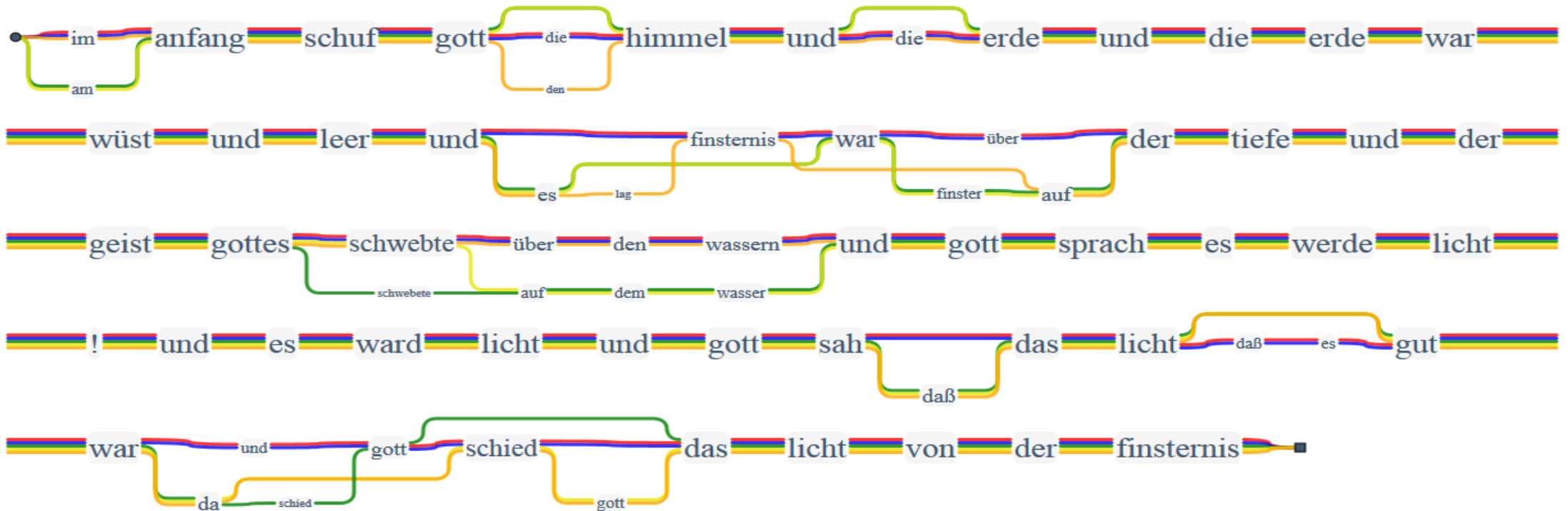
Visualisierung: [Stefan Jaenickes TRAViz](#)



# Alignierung

<http://ctstest.informatik.uni-leipzig.de/TRAViz2/test/bible/urns.html?urn=urn:cts:pb:deu.elberfelder1905:1.1.1-1.1.4>

Visualisierung: [Stefan Jaenickes TRAViz](#)



Stefan Jänicke, Leipzig University  
DEV in BMBF-project eTRACES (PN 01UA1101A)

# Generic Reader/Browser

demo - songs - farsilLit - greekLit - latinLit - pdlrefwk - dta

tlg0001 - tlg0003 - tlg0004 - tlg0005 - tlg0006 - tlg0007 - tlg0008 - tlg0010 - tlg0011 - tlg0012 - tlg0013 - tlg0014 - tlg001 - tlg002 - tlg003 - tlg004 - tlg005 - tlg006 - tlg007 - tlg008

eng2 - grc1

View Mode: ☒ TEI ☐ Plain Text ☐ Styled (experimental)

Set Citation Export Citation Show Links

9	1	nobler beasts. However, some say that Phaea was a female robber, a woman of audacious and unbridled spirit, who dwelt in Crommyon, was called Sow because of her life and manners, and was afterwards slain by Theseus.
10	2	<p>&lt;p xmlns="http://www.tei-c.org/ns/1.0"&gt; Megarian writers, however, taking issue with current report, and, as Simonides &lt;note resp="ed" place="unspecified" anchored="true"&gt; Fragment 193 (Bergk) &lt;/note&gt; expresses it, &lt;q&gt; waging war with antiquity, &lt;/q&gt; say that Sciron was neither a violent man nor a robber, but a chastiser of robbers, and a kinsman and friend of good and just men. For Aeacus, they say, is regarded as the most righteous of Hellenes, and Cychreus the Salaminian has divine honors at Athens, and the virtues of Peleus and Telamon are known to all men.</p>
11	3	<p>&lt;p xmlns="http://www.tei-c.org/ns/1.0"&gt; Well, then, Sciron was a son-in-law of Cychreus, father-in-law of Aeacus, and grandfather of Peleus and Telamon, who were the sons of Endeis, daughter of Sciron and Chariclo. It is not likely, then, they say, that the best of men made family alliances with the basest, receiving and giving the greatest and most valuable pledges. It was not, they say, when Theseus first journeyed to Athens, but afterwards, that he captured Eleusis from the Megarians, having circumvented Diocles its ruler, and slew Sciron. Such, then, are the contradictions in which these matters are involved.</p>
12	1	<p>&lt;p xmlns="http://www.tei-c.org/ns/1.0"&gt; In Eleusis, moreover, he out-wrestled Cercyon the Arcadian and killed him and going on a little farther, at Erineus, he killed Damastes, surnamed Procrustes, by compelling him to make his own body fit his bed, as he had been wont to do with those of strangers. And he did this in imitation of Heracles. For that hero punished those who offered him violence in the manner in which they had plotted to serve him, and therefore sacrificed Busiris, wrestled Antaeus to death, slew Cynus in single combat, and killed Termerus by dashing in his skull.</p> <p>&lt;p xmlns="http://www.tei-c.org/ns/1.0"&gt; It is from him, indeed, as they say, that the name &lt;q&gt; Termerian mischief &lt;/q&gt; comes, for Termerus, as it would seem, used to kill those who encountered him by dashing his head against theirs. Thus Theseus also went on his way chastising the wicked, who were visited with the same violence from him which they were visiting upon others, and suffered justice after the manner of their own injustice.</p> <p>&lt;p xmlns="http://www.tei-c.org/ns/1.0"&gt; As he went forward on his journey and came to the river Cephissus, he was met by men of the race of the Phytalidae, who greeted him first, and when he asked to be purified from bloodshed, cleansed him with the customary rites, made propitiatory sacrifices, and feasted him at their house. This was the first kindness which he met with on his journey.</p> <p>&lt;p xmlns="http://www.tei-c.org/ns/1.0"&gt; It was, then, on the eighth day of the month Cronius, now called Hecatombaeon, that he is said to have arrived at Athens. And when he entered the city, he found public affairs full of confusion and dissension,</p>

2014 Leipzig University // Martin Reckziegel

# Generic Reader TEI vs. Styled

```
<head>
| To the right honorable Sir William Cavendish, Knight of the Bath, Baron of Hardwick, and Earl of Devonshire
</head>
<p>
Right Honourable, I take confidence from your Lordship's goodness in the very entrance of this Epistle, to
profess, with simplicity and according to the faith I owe my master now in heaven, that it is not unto yourself,
but to your Lordship's father that I dedicate this my labour, such as it is. For neither am I at liberty to make
choice of one to whom I may present it as a voluntary oblation; being bound in duty to bring it in as an account
to him, by whose indulgence I had both the time and ammunition to perform it. Nor if such obligation were
removed, know I any to whom I ought to dedicate it rather. For by the experience of many years I had the honour
to serve him, I know this: there was not any, who more really, and less for glory's sake favoured those that
studied the liberal arts liberally, than my Lord your father did; nor in whose house a man should less need the
university than in his. For his own study, it was bestowed, for the most part, in that kind of learning which
best deserveth the pains and hours of great persons, history and civil knowledge: and directed not to the
ostentation of his reading, but to the government of his life and the public good. For he read, so that the
learning he took in by study, by judgment he digested, and converted into wisdom and ability to benefit his
country: to which also he applied himself with zeal, but such as took no fire either from faction or ambition.
And as he was a most able man, for soundness of advice and clear expression of himself, in matters of difficulty
and consequence, both in public and private: so also was he one whom no man was able either to draw or justle out
of the straight path of justice. Of which virtue, I know not whether he deserved more by his severity in imposing
it (as he did to his last breath) on himself, or by his magnanimity in not exacting it to himself from others. No
man better discerned of men: and therefore was he constant in his friendships, because he regarded not the
<hi>
| fortune
</hi>
nor
<hi>
| adherence,
</hi>
but the
<hi>
| men;
</hi>
with whom also he conversed with an openness of heart that had no other guard than his own integrity and that
<foreign lang="la">
| nil conscire
</foreign>
. To his equals he carried himself equally, and to his inferiors familiarly; but maintaining his respect fully,
and only with the native splendour of his worth. In sum, he was one in whom might plainly be perceived, that
<hi>
| honour
</hi>
and
<hi>
| honesty
</hi>
are but the same thing in the different degrees of persons. To him therefore, and to the memory of his worth, be
consecrated this, though unworthy, offering.
</p>
<p>
```

## To the right honorable Sir William Cavendish, Knight of the Bath, Baron of Hardwick, and Earl of Devonshire

Right Honourable, I take confidence from your Lordship's goodness in the very entrance of this Epistle, to profess, with simplicity and according to the faith I owe my master now in heaven, that it is not unto yourself, but to your Lordship's father that I dedicate this my labour, such as it is. For neither am I at liberty to make choice of one to whom I may present it as a voluntary oblation; being bound in duty to bring it in as an account to him, by whose indulgence I had both the time and ammunition to perform it. Nor if such obligation were removed, know I any to whom I ought to dedicate it rather. For by the experience of many years I had the honour to serve him, I know this: there was not any, who more really, and less for glory's sake favoured those that studied the liberal arts liberally, than my Lord your father did; nor in whose house a man should less need the university than in his. For his own study, it was bestowed, for the most part, in that kind of learning which best deserveth the pains and hours of great persons, history and civil knowledge: and directed not to the ostentation of his reading, but to the government of his life and the public good. For he read, so that the learning he took in by study, by judgment he digested, and converted into wisdom and ability to benefit his country: to which also he applied himself with zeal, but such as took no fire either from faction or ambition. And as he was a most able man, for soundness of advice and clear expression of himself, in matters of difficulty and consequence, both in public and private: so also was he one whom no man was able either to draw or justle out of the straight path of justice. Of which virtue, I know not whether he deserved more by his severity in imposing it (as he did to his last breath) on himself, or by his magnanimity in not exacting it to himself from others. No man better discerned of men: and therefore was he constant in his friendships, because he regarded not the fortune nor adherence, but the men; with whom also he conversed with an openness of heart that had no other guard than his own integrity and that nil conscire. To his equals he carried himself equally, and to his inferiors familiarly; but maintaining his respect fully, and only with the native splendour of his worth. In sum, he was one in whom might plainly be perceived, that honour and honesty are but the same thing in the different degrees of persons. To him therefore, and to the memory of his worth, be consecrated this, though unworthy, offering.

And now, imitating in this civil worship the religious worship of the gentiles; who, when they dedicated any thing to their gods, brought and presented the same to their images: I bring and present this gift of mine, THE HISTORY OF THUCYDIDES, translated into English with much more diligence than elegance, to your Lordship; who are the image of your father, (for never was a man more exactly copied out than he in you), and who have in you the seeds of his virtues already springing up: humbly intreating your Lordship to esteem it amongst the goods that descend upon you, and in your due time to read it. I could recommend the author unto you, not impertinently, for that he had in his veins the blood of kings; but I choose rather to recommend him for his writings, as having in them profitable instruction for noblemen, and such as may come to have the managing of great and weighty actions. For I may confidently say, that notwithstanding the excellent both examples and precepts of heroic virtue you have at home, this book will confer not a little to your institution; especially when you come to the years to frame your life by your own observation. For in history, actions of honour and dishonour do appear plainly and distinctly, which are which; but in the present age they are so disguised, that few there be, and those very careful, that be not grossly mistaken in them. But this, I doubt not, is superfluously spoken

Copyright 2014 Leipzig University // Martin Reckziegel

2014 Leipzig University // Martin Reckziegel

# CTS – Text Miner (CTSTM)

## CTS Text Mining Framework

**Broad and comprehensive framework for text analysis**

**Done:**

**Term-Document Matrix**

Token/Types per Document/Corpus

**Document- and Termbased Pruning + lists of  
Stopwords**

**Tokensequence /(Kookurenz)**

**-> Fulltextsearch -> Citation Analysis ....**

# CTS – Additional Functions

**More request possibilities without XML documents**

[Editions](#)

[Authors](#)

[CompactDepthTypes](#)

[DepthTypes](#)

[AlignmentCandidates](#)

[Alignment \(Bible ind\)](#)

[Alignment \(Bible de\)](#)

...

# CTS Admin Tool

CTS Admin Tool

create new CTSupdate CTS templatecontact J. Tiepmar

signed in as **cts**logout

list of all CTS:

demo

dta2

law

pbc

perseus

DB-Config

Data Import

Servlet

Browse Data

divs	<input type="checkbox"/>
epidoc	<input type="checkbox"/>
escapePassage	<input checked="" type="checkbox"/>
maxlevelexception	<input type="checkbox"/>
seperatecontext	<input checked="" type="checkbox"/>
stats	<input type="checkbox"/>
xmiformating	<input checked="" type="checkbox"/>
smallinventory	<input type="checkbox"/>

save parameters

rename this CTS instance ("demo")

The renaming of an CTS instance can take up to 30 seconds. Please stand by and do not leave this page. A message will appear when the CTS instance has been renamed successfully. For the CTS instance name only small and uppercase letters as well as numbers and the underscore are available ( a-z A-Z 0-9 \_ ). All other characters will be removed automatically. Thus, any white-spaces and special characters will be removed.

new name:

new name

rename

delete this CTS instance ("demo")

The deletion of an CTS instance can take up to 30 seconds. Please stand by and do not leave this page. A message will appear when the CTS instance has been deleted successfully. Be careful! The deletion of an CTS instance cannot be undone!

type uppercase OK as confirmation

confirm with OK

delete (cannot be undone!)

Implemented by  
Sascha Ludwig

# CTS Cloning

URNs specify @n-Value of <div>s

-> @n-Values can be used to reconstruct URNs

-> Content of one CTS can be cloned

Data can be narrowed down „from left to right“ by URNs

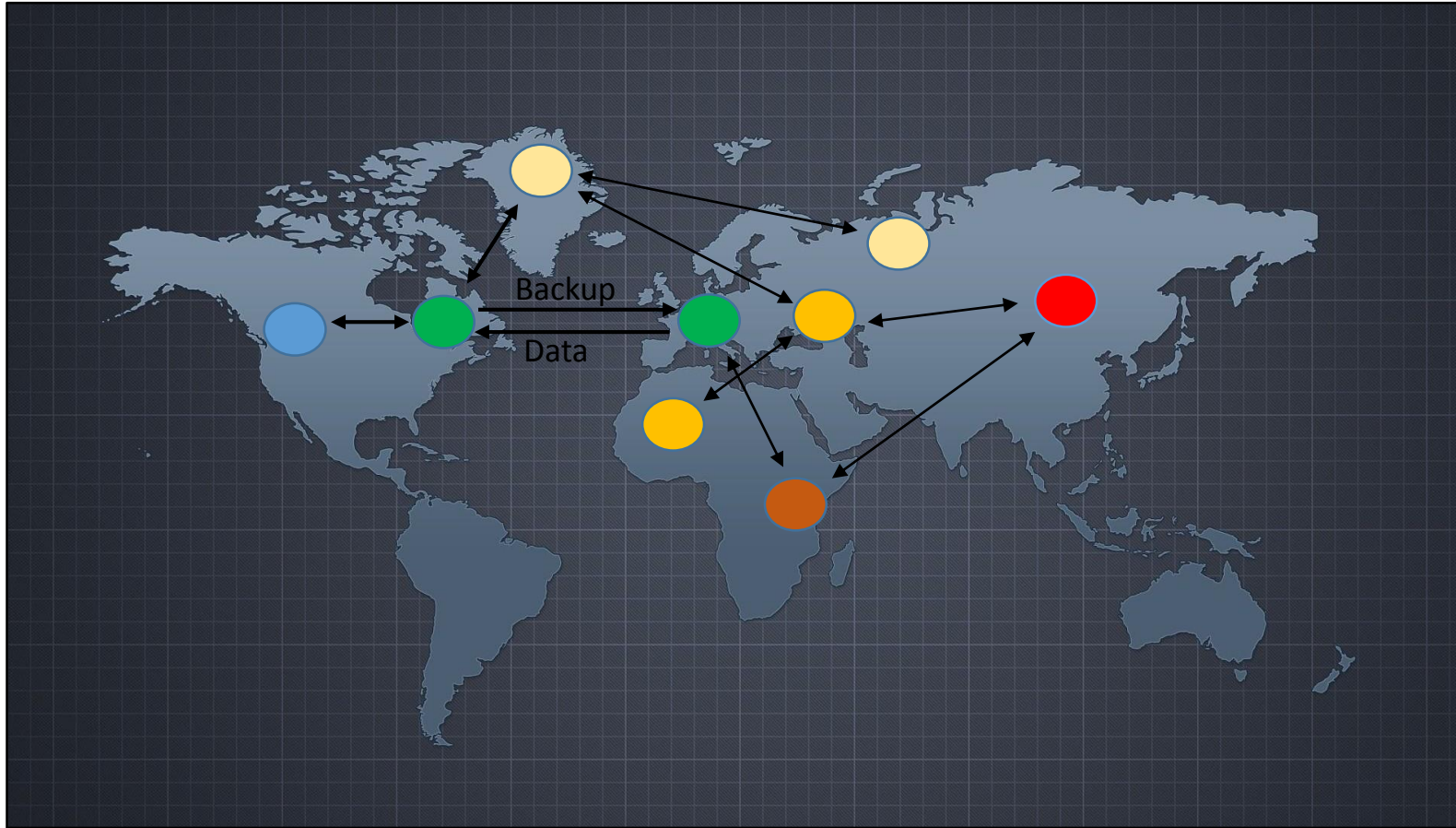
Clone everything from Shakespeare:

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

```
<passage>
  <div1 n="1" type="song">
    <div2 n="1" type="strophe">
      <div3 n="1" type="line">
      </div3>
    </div2>
    <div2 n="2" type="strophe">
      <div3 n="1" type="line">
      </div3>
    </div2>
  </div1>
</passage>
```

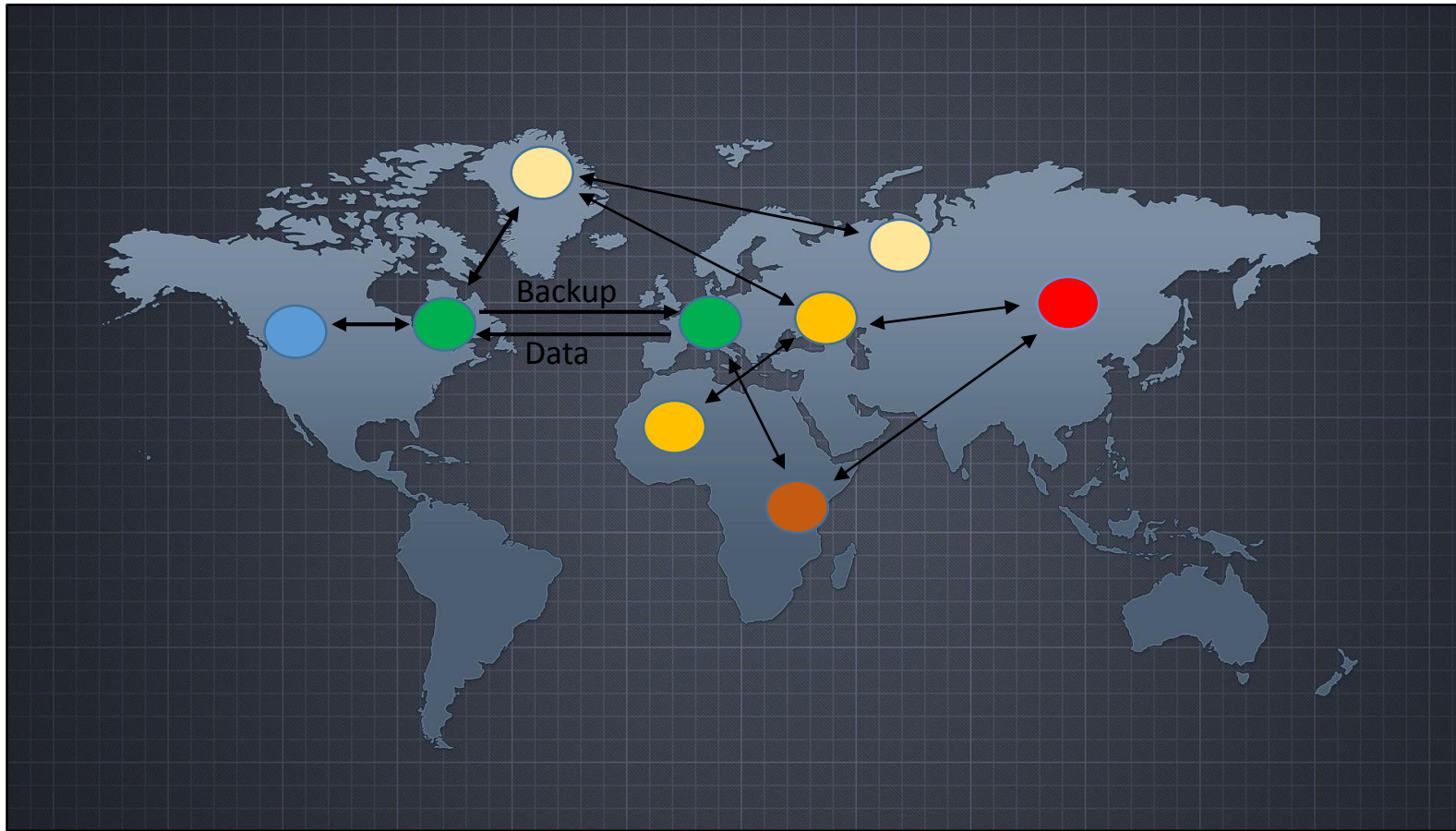


# CTS Cloning



<http://hdw.eweb4.com/out/1369880.html>

# Big Picture



**global  
decentralised  
community organised  
community backup'ed  
open access  
standardized  
persistent citable  
easy to install  
text repository  
for browsing, searching  
and analysis of text resources.**

<http://hdw.eweb4.com/out/1369880.html>