

Canonical Text Service

Jochen Tiepmar

ScaDS, ASV, Uni Leipzig



Federal Ministry
of Education
and Research

Overview CTS

Canonical Text Services (CTS)

- protocol for a webbased citable text service
- Unique Identifiers (**Unique Resource Name, URN**) refer to text passages
- Developed in Homer Multitext Project (www.homermultitext.org), Smith et.al. 2009
<http://www.homermultitext.org/hmt-docs/specifications/ctsurn/>
<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

Text as web resource

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

SPIEGEL ONLINE KULTUR

Politik | Wirtschaft | Panorama | Sport | Kultur | Netzwerk | 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

Navigation: Kapitel 2

Erster Aufzug.
Erste Scene.
(Eine Terrasse vor dem Palast.)
(Bernardo und Francisco, zweien 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>

Included in project specific website

No –obvious- possibility for automated processing

Crawling can be problematic (request limit, advertisement, Javascript,...)

Hamlet by von Schlegel

```
<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 by Wieland

```
<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 by Moby

```
<h3>SCENE I. Elsinore. A platform before the castle.</H3>
<p><blockquote>
<i>FRANCISCO at his post. Enter to him BERNARDO</i>
</blockquote>
<A NAME=speech1><b>BERNARDO</b></a>
<blockquote>
<A NAME=1>Who's there?</A><br>
</blockquote>
<A NAME=speech2><b>FRANCISCO</b></a>
<blockquote>
<A NAME=2>Nay, answer me: stand, and unfold yourself.</A><br>
</blockquote>
<A NAME=speech3><b>BERNARDO</b></a>
<blockquote>
<A NAME=3>Long live the king!</A><br>
</blockquote>
```

Hamlet by 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="ftIn" xml:id="ftIn-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>
```

Variety

No commonly used text format & access

No distinction between document structure (chapter), meta information (author) and text

Files are incompatible to other project workflows

TEI/XML

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

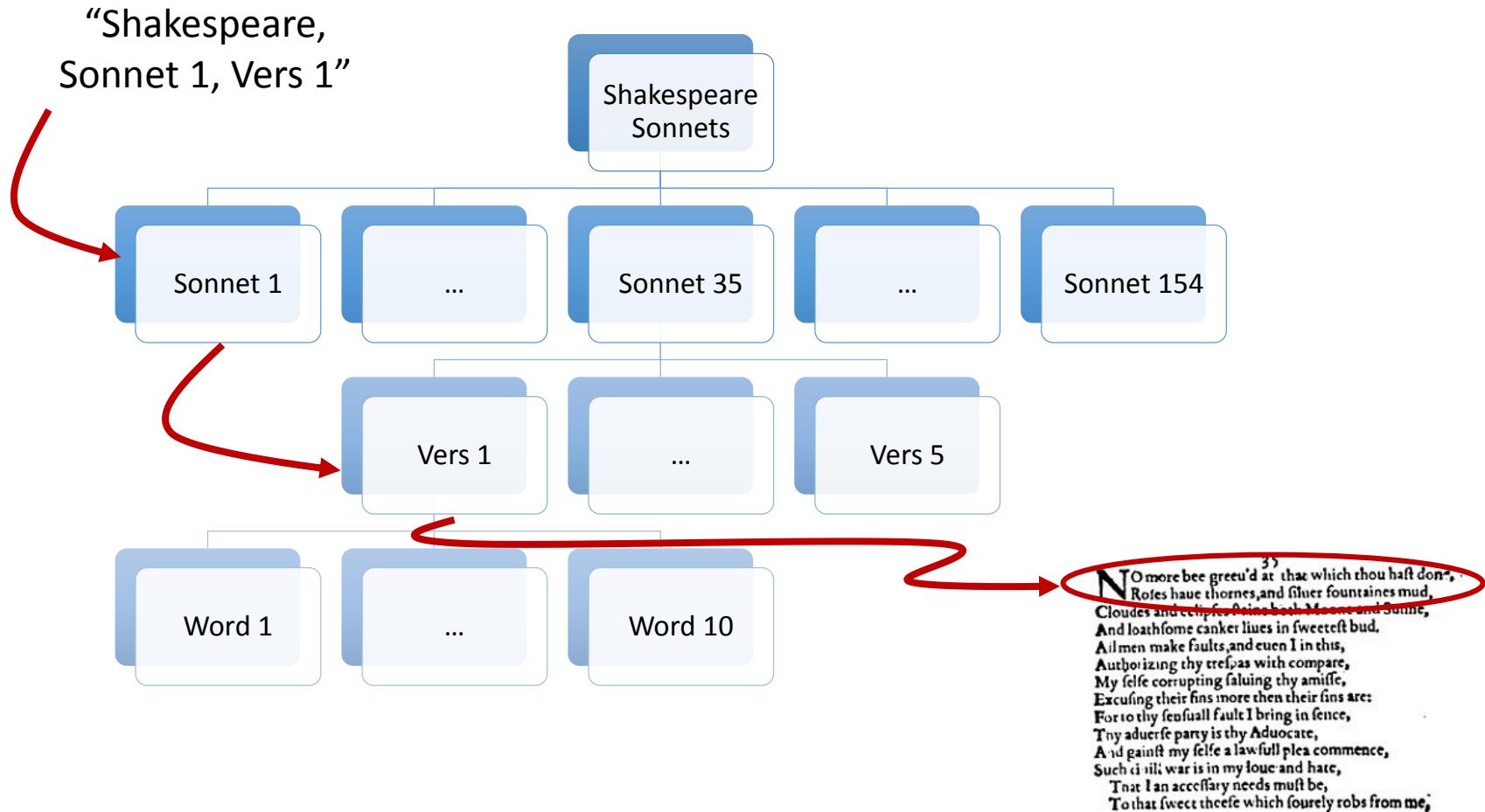
```
<teiheader> ... </teiheader>
<text id="stoa0045.stoa0"><body>
<div1 n="Sh." type="Book">
< milestone ed="P" n="1" unit="card"/>
<!--
< milestone ed="P" unit="para"/>ἢ οἵ προλιποῦσα δόμους καὶ πατρίδα γαῖαν</l>
<l>ἢ λυθεν ἐς Θήβας μετ' ἀρήιον Ἀμφιτρύωνα</l>
<l>Ἀλκμήνη, θυγάτηρ λαοσσόου Ἡλεκτρύωνος·</l>
<l>ἢ ρα γυναικῶν φῦλον ἐκαίνυτο θηλυτεράων</l>
<l n="5">εἴδεῖ τε μεγέθει τε· νόον γε μὲν οὕ τις ἔριζε</l>
<l>τάων, ἃς θνηταὶ θνητοῖς τέκον εύνηθεῖσαι.</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,&mdash;even Alcmena, the daughter of Electryon, gatherer of the people. She surpassed the tribe of 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 mortal men. Her face and her dark eyes wafted 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 before her. In truth he had slain her noble father 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 dwelt with his modest wife< 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-hearted brothers and utterly burned with blazing fire 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 it. And he feared their anger, and hastened 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"/>and the Locrians who fight hand to 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"/>So he arose from <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. Quickly he came to Typhaonium, and from there 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> which took its name from 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 marvellous things in his heart.< 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, gatherer of the people, the glorious hero, came to his house when he had ended his great task. </p>
</div2>
```

Documents Hierarchy



Citation

Document „outer hierarchy“

Shakespeare → Sonnets → english → 1st edition

Text passage „inner hierarchy“

Sonnet 1 → Vers 1

Combined

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

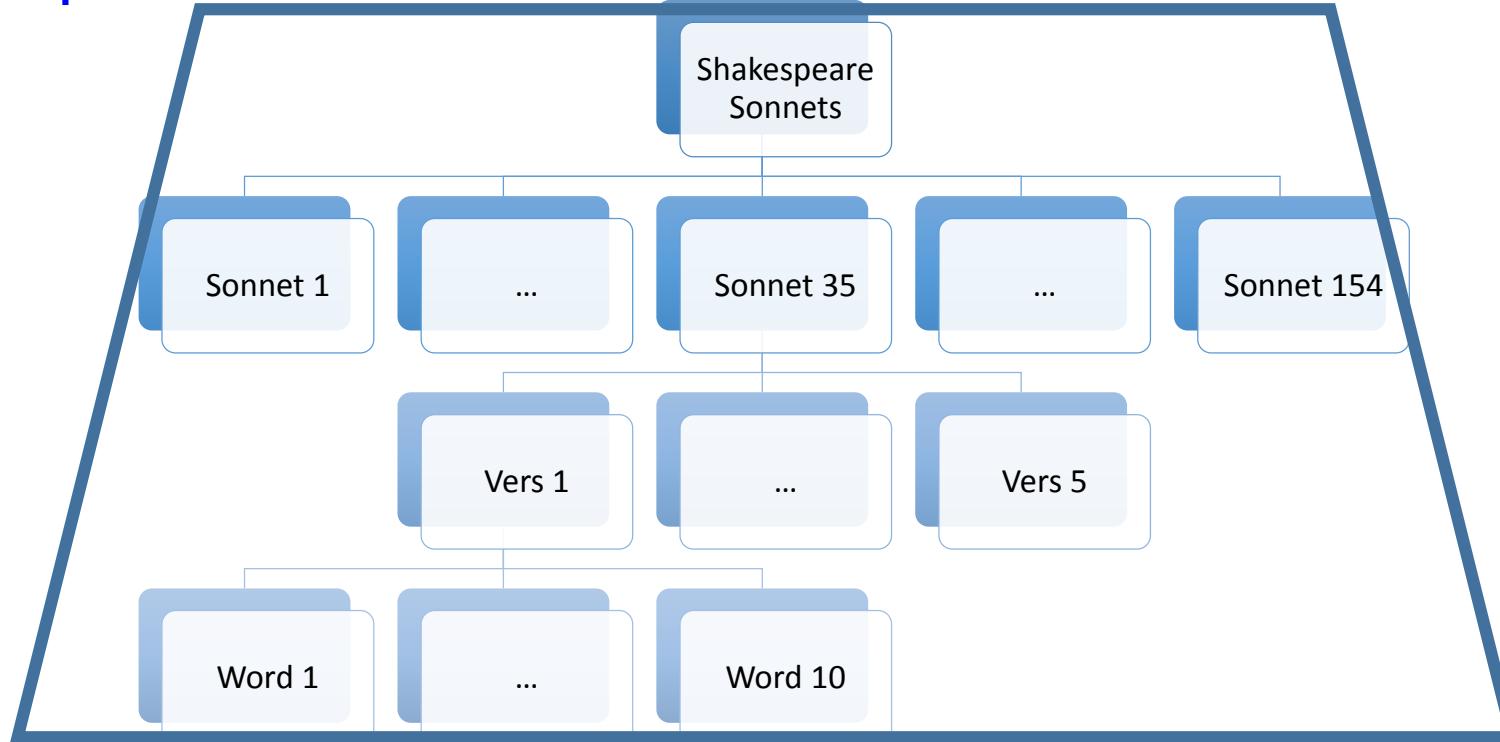
CTS-URN

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

Canonical Citation

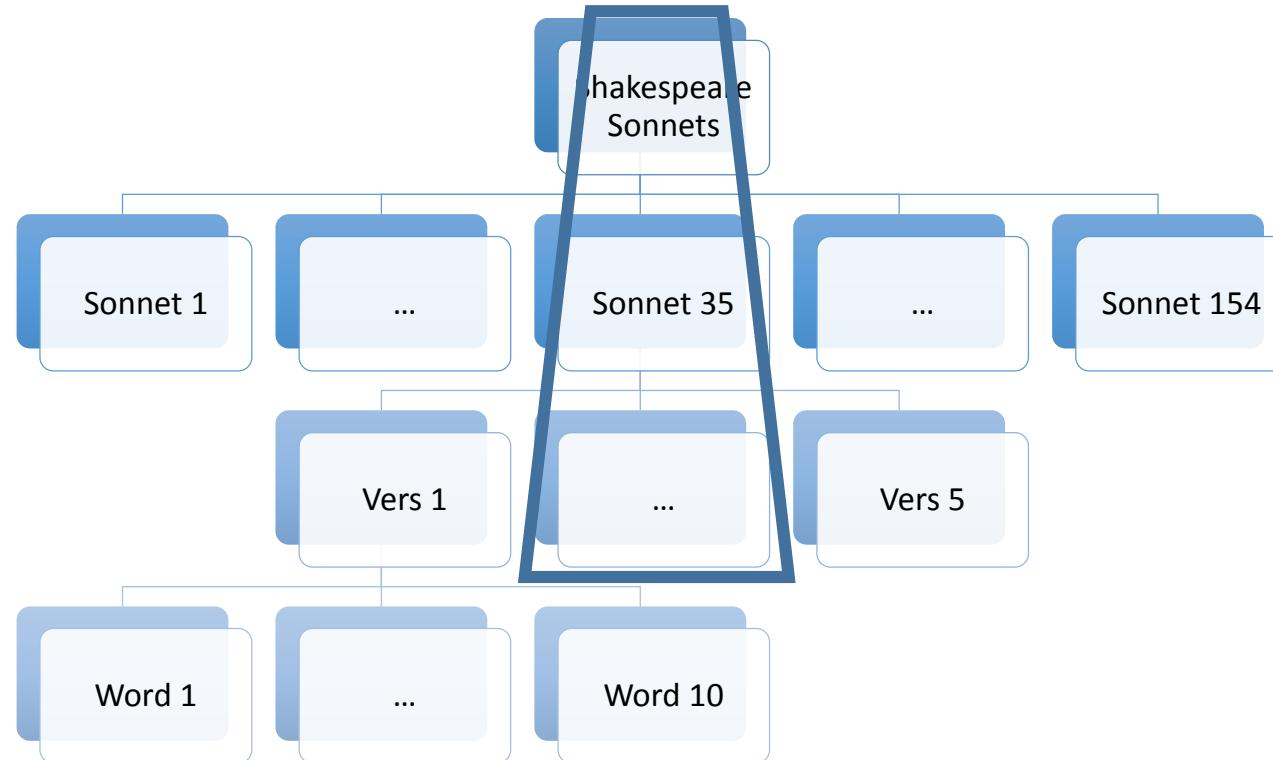
urn:cts:demo:shakespeare.sonnets:

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



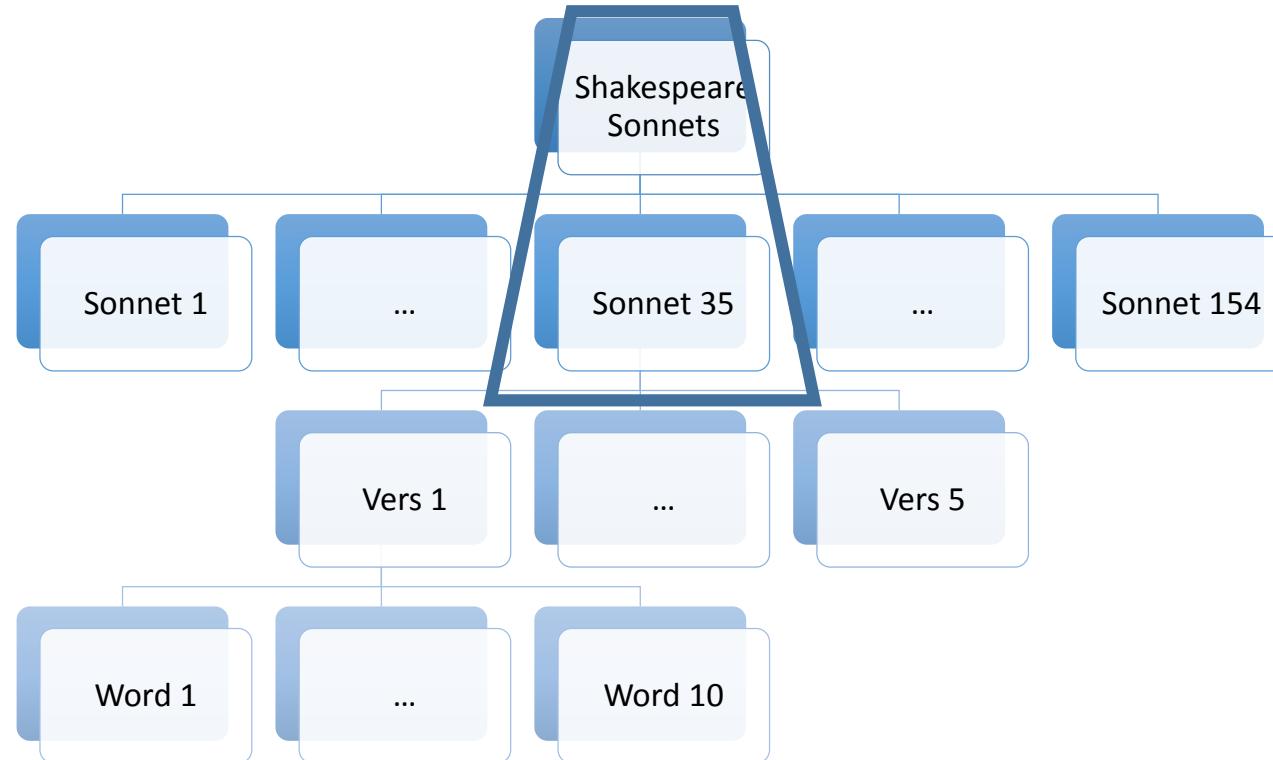
Canonical Citation

urn:cts:demo:shakespeare.sonnets:35.4



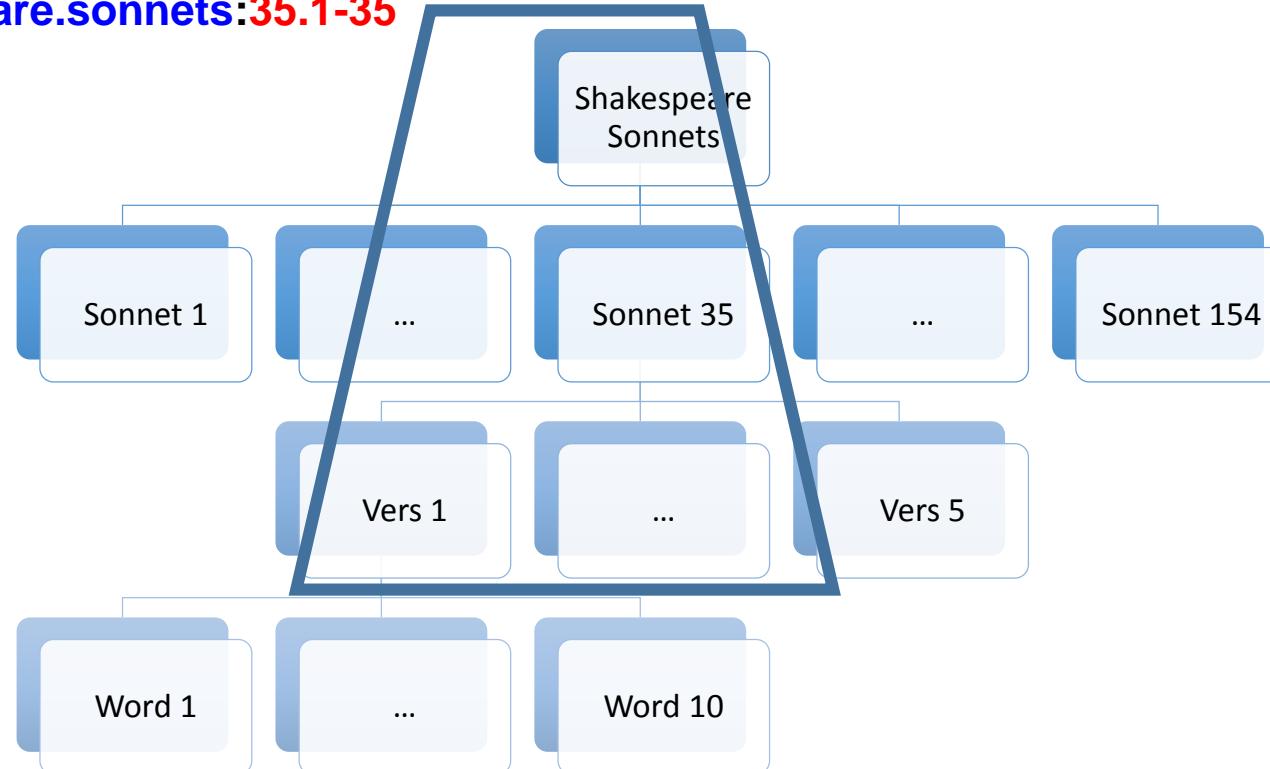
Canonical Citation

urn:cts:demo:shakespeare.sonnets:35



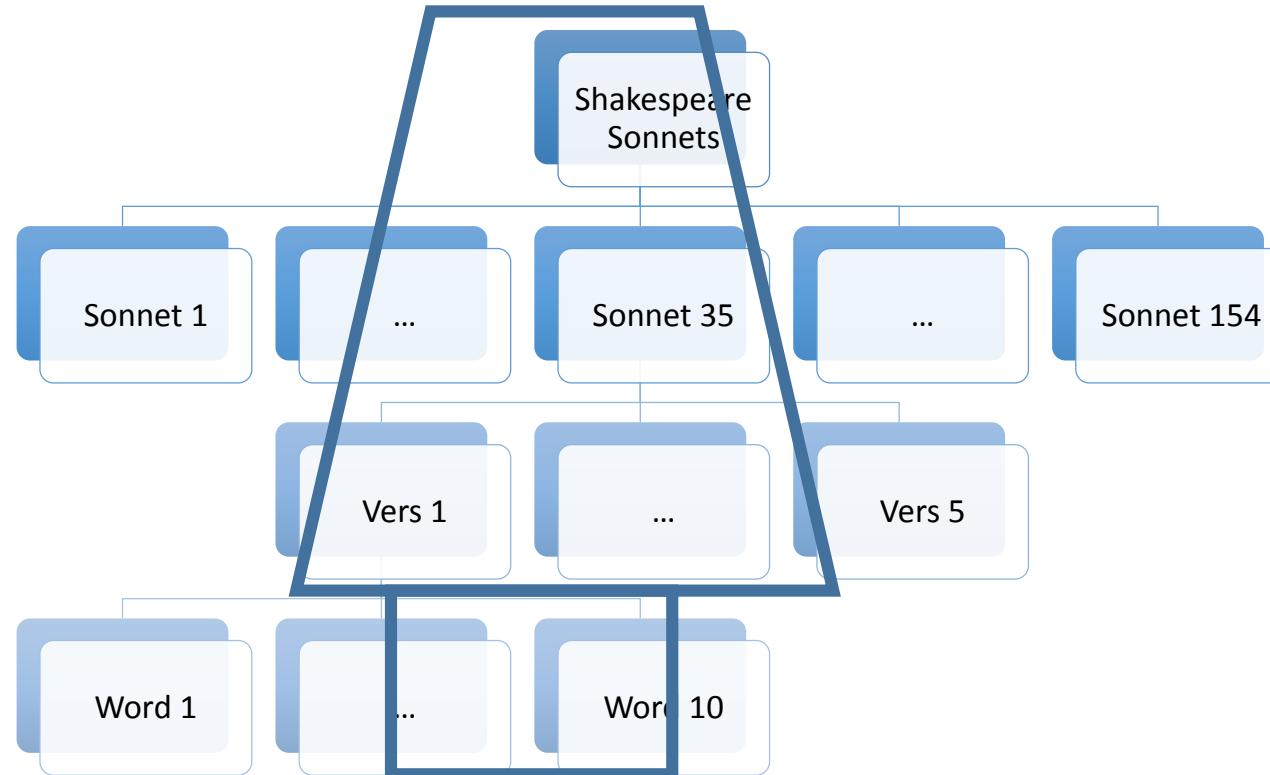
Canonical Citation

urn:cts:demo:shakespeare.sonnets:35.1-35.5
urn:cts:demo:shakespeare.sonnets:35.1-35



Canonical Citation

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



CTS Requests

<http://ctstest.informatik.uni-leipzig.de/perseus/cts/?request=GetPassage&urn=urn:cts:latinLit:phi1014.phi001.lat1:1.pr@magis-1.2.1@est>

CTS Requests

<http://ctstest.informatik.uni-leipzig.de/perseus/cts/?request=GetPassage&urn=urn:cts:latinLit:phi1014.phi001.lat1:1.pr@magis-1.2.1@est>

Server, Endpoint, Access Point

CTS Requests

<http://ctstest.informatik.uni-leipzig.de/perseus/cts/?request=GetPassage&urn=urn:cts:latinLit:phi1014.phi001.lat1:1.pr@magis-1.2.1@est>

Request

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

CTS Requests

<http://ctstest.informatik.uni-leipzig.de/perseus/cts/?request=GetPassage&urn=urn:cts:latinLit:phi1014.phi001.lat1:1.pr@matis-1.2.1@est>

Parameters for request

URN, (citation)level,...

CTS Requests

: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.

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>
```

Using CTS to standardize texts

Differentiate text structure from text content and meta information

Refer to generic text parts

Reduce type of text part to label

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

XML Styled

Sophocles
Trachiniae (English) Machine readable text

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 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, is he on the straits of the sea, or does he lean upon the twin continents? Speak, you who surpass all in sight!

Chorus
 For with longing heart, as I hear, Deianeira, the battle-prize, now like some mournful bird, never rests her eyes' longing that they might be without tears, but nourishing a well-remembered fear for her husband's travels she is constantly afflicted 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 dissent. You must not, I say, wear away fair hope.

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

XML Styled

Sophocles
Trachiniae (English) Machine readable text

```

</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">
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>

```

Copyright 2014 Leipzig University // Martin Reckziegel

2014 Leipzig University // Martin Reckziegel

CTS Cloning

URNs specify @n-Value of <div>s

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

-> Content of one CTS can be cloned

URNs can be used to filter data

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>

Alignment

Find comparable text parts in different documents

-> Comparison of documents/editions

-> Input for translation memory tools

Alignment

Naiv: Sentence tokenisation

-> Sentences can be splitted/joined

Named Entity Recognition

-> Statistical methods

Pre-calculation, Cross Comparison

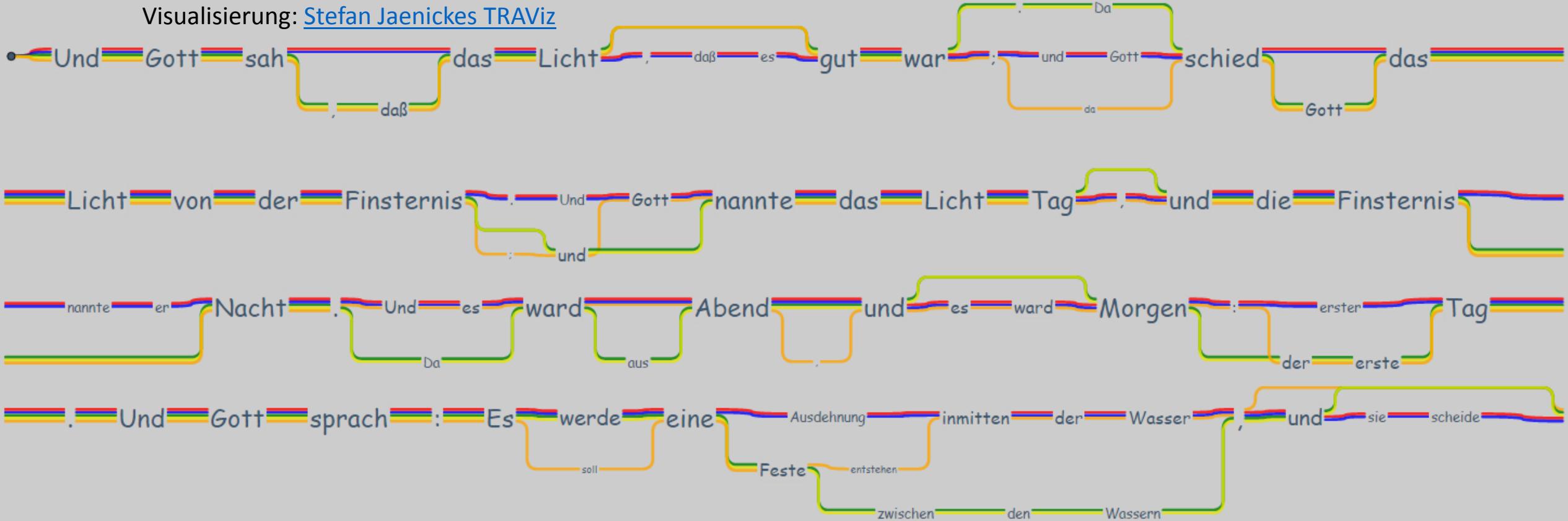
-> DTA: 3.5 GB * 5136 documents > 17 TB

Error probability increases with each text unit

Alignment

<http://ctstest.informatik.uni-leipzig.de/Alignment/?cts=pbc&urn=urn:cts:pbc:deu.luther1545:1.1.4-1.1.7>

Visualisierung: [Stefan Jaenickes TRAViz](#)



Alignment via CTS

<http://ctstest.informatik.uni-leipzig.de/demo/plain/alignment?urn=urn:cts:songs:christmas.ohtennenbaum.de.1:&alignurns=urn:cts:songs:christmas.ohtennenbaum.en.1: urn:cts:songs:christmas.ohtennenbaum.lat.1:>

section	urn:cts:songs:christmas.ohtennenbaum.de.1:	urn:cts:songs:christmas.ohtennenbaum.en.1:	urn:cts:songs:christmas.ohtennenbaum.lat.1:
:1			
:1.1			
:1.1.1	O Tannenbaum, O Tannenbaum,	O Christmas tree, O Christmas tree !	O abies, o abies,
:1.1.2	Wie treu sind deine Blätter.	How are thy leaves so verdant !	Fidissima cum fronde,
:1.1.3	Du grünst nicht nur zur Sommerzeit,	Not only in the summertime,	Virens æstatis tempore,
:1.1.4	Nein auch im Winter wenn es schneit.	But even in winter is thy prime.	Virens nivosa hieme,
:1.1.5	O Tannenbaum, O Tannenbaum,	O Christmas tree, O Christmas tree,	O abies, o abies,
:1.1.6	Wie grün sind deine Blätter!	How are thy leaves so verdant !	Fidissima cum fronde !
:1.2			
:1.2.1	O Tannenbaum, O Tannenbaum,	O Christmas tree, O Christmas tree,	O abies, o abies,
:1.2.2	Du kannst mir sehr gefallen!	Much pleasure doth thou bring me !	Tu mihi valde places :
:1.2.3	Wie oft hat schon zur Winterszeit	For every year the Christmas tree,	Quotiens natali tempore
:1.2.4	Ein Baum von dir mich hoch erfreut!	Brings to us all both joy and glee.	Delectavisti me fronde,
:1.2.5	O Tannenbaum, O Tannenbaum,	O Christmas tree, O Christmas tree,	O abies, o abies,
:1.2.6	Du kannst mir sehr gefallen!	Much pleasure doth thou bring me !	Tu mihi valde places.
:1.3			
:1.3.1	O Tannenbaum, O Tannenbaum,	O Christmas tree, O Christmas tree,	O abies, o abies,
:1.3.2	Dein Kleid will mich was lehren:	Thy candles shine out brightly !	Veste tua me doces :
:1.3.3	Die Hoffnung und Beständigkeit	Each bough doth hold its tiny light,	Spes gignunt et constantia
:1.3.4	Gibt Mut und Kraft zu jeder Zeit!	That makes each toy to sparkle bright.	Semper vires, solacia ;
:1.3.5	O Tannenbaum, O Tannenbaum,	O Christmas tree, O Christmas tree,	O abies, o abies,
:1.3.6	Dein Kleid will mich was lehren.	Thy candles shine out brightly !	Veste tu me doces

Alignment via CTS

Alignment of URNs

urn:cts:demo:shakespeare.sonnets.en.1:3.2

urn:cts:demo:shakespeare.sonnets.de.1:3.2

urn:cts:demo:shakespeare.sonnets.fr.1:3.2

-> Error probability resets for each new chapter

-> fast & real time, no pre-calculation

...paper incoming...

section	urn:cts:songs:christmas.ohthennenbaum.de.1:	urn:cts:songs:christmas.ohthennenbaum.en.1:	urn:cts:songs:christmas.ohthennenbaum.fr.1:
:1			
:1.1			
:1.1.1	O Tannenbaum, O Tannenbaum,	O Christmas tree, O Christmas tree !	O abies, o abies,
:1.1.2	Wie treu sind deine Blätter.	How are thy leaves so verdant !	Fidissima cum fronde,
:1.1.3	Du grünst nicht nur zur Sommerzeit,	Not only in the summertime,	Virens aestatis tempore,
:1.1.4	Nein auch im Winter wenn es schneit.	But even in winter is thy prime.	Virens nivosa hieme,
:1.1.5	O Tannenbaum, O Tannenbaum,	O Christmas tree, O Christmas tree,	O abies, o abies,
:1.1.6	Wie grün sind deine Blätter!	How are thy leaves so verdant !	Fidissima cum fronde !
:1.2			
:1.2.1	O Tannenbaum, O Tannenbaum,	O Christmas tree, O Christmas tree,	O abies, o abies,
:1.2.2	Du kannst mir sehr gefallen!	Much pleasure doth thou bring me !	Tu mihi valde places :
:1.2.3	Wie oft hat schon zur Winterszeit	For every year the Christmas tree,	Quotiens natali tempore
:1.2.4	Ein Baum von dir mich hoch erfreut!	Brings to us all both joy and glee.	Delectavisti me fronde,
:1.2.5	O Tannenbaum, O Tannenbaum,	O Christmas tree, O Christmas tree,	O abies, o abies,
:1.2.6	Du kannst mir sehr gefallen!	Much pleasure doth thou bring me !	Tu mihi valde places.
:1.3			
:1.3.1	O Tannenbaum, O Tannenbaum,	O Christmas tree, O Christmas tree,	O abies, o abies,
:1.3.2	Dein Kleid will mich was lehren:	Thy candles shine out brightly !	Veste tua me doces :
:1.3.3	Die Hoffnung und Beständigkeit	Each bough doth hold its tiny light,	Spes gignunt et constantia
:1.3.4	Gibt Mut und Kraft zu jeder Zeit!	That makes each toy to sparkle bright.	Semper vires, solacia ;
:1.3.5	O Tannenbaum, O Tannenbaum,	O Christmas tree, O Christmas tree,	O abies, o abies,
:1.3.6	Dein Kleid will mich was lehren.	Thy candles shine out brightly !	Veste tu me doces

Alignment via CTS

<http://ctstest.informatik.uni-leipzig.de/pbc/plain/alignment?urn=urn:cts:pbc:ben.common:1.4-1.5&alignurns=urn:cts:pbc:fra.kingjames::urn:cts:pbc:eng.darby::urn:cts:pbc:deu.luther1912;>

:1.4.21	যাবলের ভাইয়ের নাম ছিল যুবল। মাঝা বীণা ও বাঁশি বাজায় যুবল তাদের সকলের পূর্বৰূপ।	Et le nom de son frère était Jubal ; il fut père de tous ceux qui jouent de la harpe et du chalumeau .	And his brother's name was Jubal : he was the father of those who handle the harp and pipe .	Und sein Bruder hieß Jubal ; von dem sind hergekommen die Geiger und Pfeifer .
:1.4.22	সিল্লার গর্তে তৃবল-কমিজের তাঙ্গা হল। প্রোজ আর সোহার সব রকমের যত্নগুচ্ছ তৈরী করা ছিল তার কাজ। তৃবল-কমিজের বোলের নাম ছিল নাম্মা।	Et Tsilla , elle aussi , enfanta Tubal-Cain , qui forgeait toute sorte d'instruments tranchants d'airain et de fer ; et la sœur de Tubal-Cain fut Naama .	And Zillah , she also bore Tubal-Cain , the forger of every kind of tool of brass and iron . And the sister of Tubal-Cain was Naamah .	Die Zilla aber gebaer auch , nämlich den Thubalkain , den Meister in allerlei Erz- und Eisenwerk . Und die Schwester des Thubalkain war Naema .
:1.4.23	একদিন লেখক তাঁর দুই স্ত্রীকে বললেন, “আম আর সিল্লা, তোমরা আমার কথা শোন ; লেমেকের স্ত্রী, আমার কথায় কান দাও ! যে লেক আমাকে জখম করছে, অর্থাৎ সে যুবক আমার গায়ে হাত তুলেছে, আমি তাকে খুন করেছি ।”	Et Lémech dit à ses femmes : Ada et Tsilla , écoutez ma voix ; femmes de Lémech , prétez l'oreille à ma parole : Oui ! j'ai tué un homme pour ma blessure , et un jeune homme pour ma meurtrissure .	And Lemech said to his wives : Adah and Zillah , hear my voice ; Ye wives of Lemech , listen to my speech . For I have slain a man for my wound , and a youth for my bruise .	Und Lamech sprach zu seinen Weibern Ada und Zilla : Ihr Weiber Lamechs , hört meine Rede und merkt , was ich sage : Ich habe einen Mann erschlagen für meine Wunde und einen Jüngling für meine Beule ;
:1.4.24	কাইনকে খুন করবার প্রতিশ্রূত যদি সাতগুণ হয়, তবে (লেখককে খুন করবার প্রতিশ্রূত হবে সাতাত্ত্ব গুণ। ”	Car si Caïn est vengé sept fois , Lémech le sera soixante-dix-sept fois .	If Cain shall be avenged seven-fold ; Lemech seventy and seven fold .	Kain soll siebenmal gerächt werden , aber Lamech siebenundsiebzigmal .
:1.4.25	পরে আদম আবার তাঁর স্ত্রীর কাছে গেলেন এবং তাঁর স্ত্রীর একটি ছেলে হল। তাঁর স্ত্রী তাঁর নাম রাখলেন ইশেঁ। হ্যাঁ বললেন, “কাইন হেবগাকে খুন করার ক্ষেত্রে হেবগার হেবেরের আগ্রাম আমাকে আর একাটি সন্তান দিলেন। ”	Et Adam connut encore sa femme , et elle enfanta un fils et l'appela Seth (remplaçant) ; car Dieu , dit-elle , m'a donné un autre fils au lieu d' Abel , que Caïn a tué .	And Adam knew his wife again , and she bore a son , and called his name Seth : . . . For God has appointed me another seed instead of Abel , because Cain has slain him .	Adam erkannte abermals sein Weib , und sie gebar einen Sohn , den hieß sie Seth ; denn Gott hat mir , sprach sie , einen andern Samen gesetzt für Abel , den Kain erwürgt hat .
:1.4.26	পরে শেখের একাটি ছেলে হল। তিনি তাঁর নাম রাখলেন ইনোশ। সেই সময় থেকে লোকেরা সন্দাপ্রভূক তাঁর যোগ সম্মান দিতে শুরু করল।	Et un fils naquit aussi à Seth , et il l'appela Énoch . Alors on commença à invoquer le nom de l' Eternel .	And to Seth , to him also was born a son ; and he called his name Enosh . Then people began to call on the name of Jehovah .	Und Seth zeugte auch einen Sohn und hieß ihn Enos . Zu der Zeit fing man an , zu predigen von des HERRN Namen .
:1.5				
:1.5.1	এই হল আদমের বংশের কথা। মানুষ সৃষ্টি করবার সময় সৈন্ধব তাঁকে তাঁর নিজের মত করে সৃষ্টি করলেন ;	Voici le livre de la postérité d' Adam . Au jour où Dieu créa l'homme , il le fit à la ressemblance de Dieu .	This is the book of Adam's generations . In the day that God created man , in the likeness of God made he him .	Dies ist das Buch von des Menschen Geschlecht . Da Gott den Menschen schuf , machte er ihn nach dem Bilde Gottes ;
:1.5.2	সৃষ্টি করলেন পূর্বে এবং স্ত্রীলোক করে এবং তাঁকের আশীর্বাদ করলেন। সৃষ্টির সময়ে তিনি তাদের নাম দিলেন মানুষ।	Il les créa mâle et femelle , et il les bénit , et leur donna le nom d' Homme , au jour qu'ils furent créés .	Male and female created he them ; and blessed them , and called their name Adam , in the day when they were created .	und schuf sie einen Mann und ein Weib und segnete sie und hieß ihren Namen Mensch zur Zeit , da sie geschaffen wurden
:1.5.3	একশো ট্রিশ বছর বয়সে আদমের একাটি ছেলের জন্ম হল। ছেলেটি বাইরে এবং ডিতে তাঁরই মত হয়েছিল। আদম তাঁর নাম দিলেন শেখ।	Or , Adam vécut cent trente ans , et engendra un fils à sa ressemblance , selon son image , et il lui donna le nom de Seth .	And Adam lived a hundred and thirty years , and begot [a son] in his likeness , after his image , and called his name Seth .	Und Adam war hundertunddreißig Jahre alt und zeugte einen Sohn , der seinem Bild ähnlich war und hieß ihn Seth
:1.5.4	শেখের তাঁরের পর আদম আরও আট্টো বছর বৈচে ছিলেন। এর মধ্যে তাঁর আরও ছালেমানের জন্ম হয়েছিল।	Et les jours d' Adam , après qu'il eut engendré Seth , furent de huit cents ans ; et il engendra des fils et des filles .	And the days of Adam after he had begotten Seth were eight hundred years ; and he begot sons and daughters .	und lebte darnach achthundert Jahre und zeugte Söhne und Töchter ;
:1.5.5	মোট ৮’ শো ট্রিশ বছর (বৈচে থাকবার পর আদম মাঝে গেলেন।	Tout le temps qu' Adam vécut , fut donc de neuf cent trente ans ; puis il mourut .	And all the days of Adam that he lived were nine hundred and thirty years ; and he died .	daß sein ganzes Alter ward neuhundertunddreißig Jahre , und starb .
:1.5.6	শেখের একশো পাঁচ বছর বয়সে তাঁর ছেলে ইনোশের জন্ম হল।	Seth vécut cent cinq ans , et engendra Énoch .	And Seth lived a hundred and five years , and begot Enos .	Seth war hundertundfünf Jahre alt und zeugte Enos

Topic Models

Find abstract topics/thematic structure in text corpus

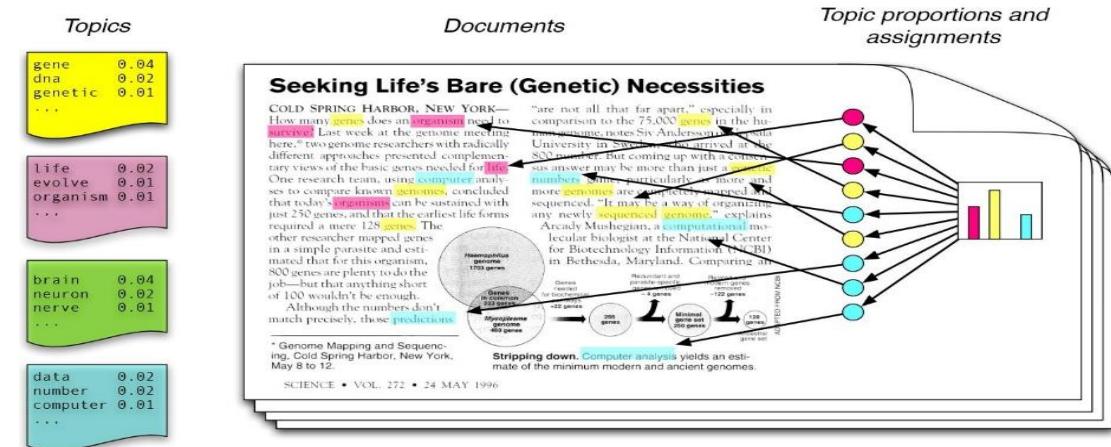
Easiest Algorithm:

Latent Dirichlet Allocation
(LDA)

Latent:

Documents share prepared set of topics
word/document ratio differs
(90% Evolution, 10% Disease)

Genetics	Evolution	Disease	Computers
Human	Evolution	Disease	Computer
Genome	Species	Host	Model
DNA	Organism	Bacteria	information



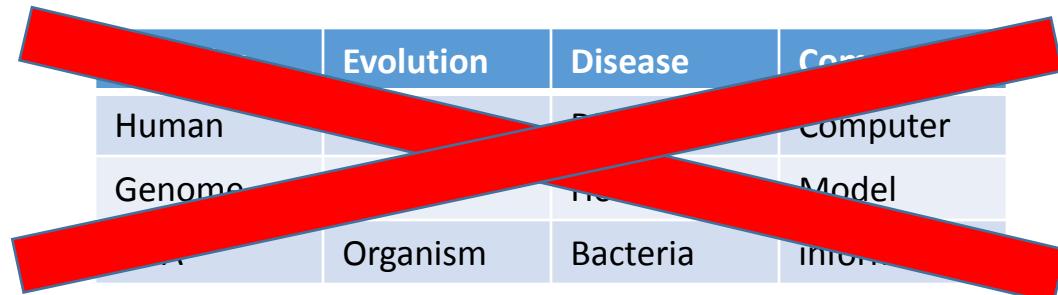
(David M. Blei: Introduction to Probabilistic Topic Models
<http://www.cs.princeton.edu/~blei/papers/Blei2011.pdf>)

Topic Models

What if the topics are unknown?

Topics can be generated with probability for common occurrence

labor	jury
workers	trial
employees	crime
union	defendant
employer	sentencing
work	judges
job	punishment
bargaining	evidence
unions	sentence
collective	offense

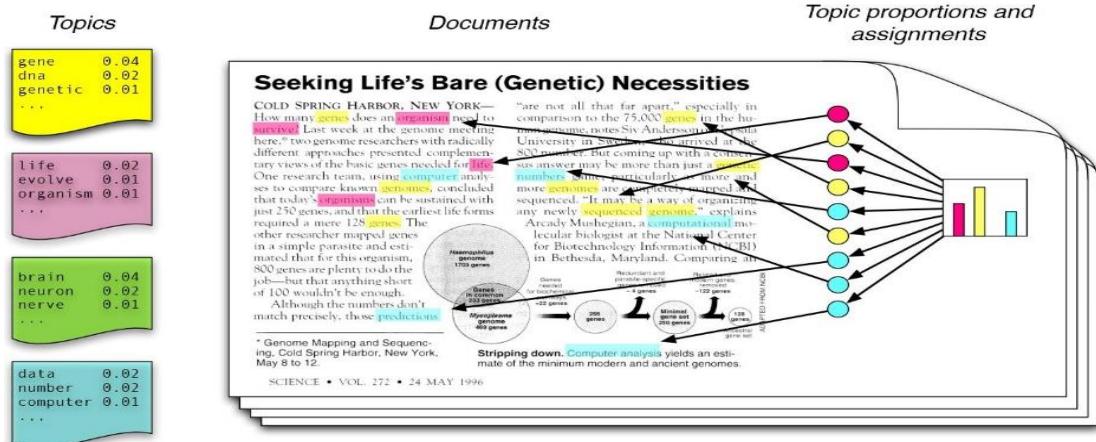


Topic Models

- Manage and order unknown text corpus
- Help explore unknown text collections
 - Search engine that lets you “zoom in or out” of topics
 - Reveal different or further aspects of one topic

Find semantic relations between texts
 Use “topic-fingerprint” of the user to propose documents

Genetics	Evolution	Disease	Computers
Human	Evolution	Disease	Computer
Genome	Species	Host	Model
DNA	Organism	Bacteria	Information



Topic Models

- Data: TED Subtitle Corpus, 51770 documents, [105 languages](#), 1938 english documents, big variety of topics
- [Tokens for Topic 25](#)

25 -> brain, and, neurons, brains, memory, consciousness, autism, cells, mind, sleep
- [Texts about topic 25](#)

„Re-engineering the brain“, „Parkinson's, depression and the switch that might turn them off“, „A prosthetic eye to treat blindness“, „How your brain tells you where you are“, „The mysterious workings of the adolescent brain“, „A monkey that controls a robot with its thoughts. No, really.“, „Your brain is more than a bag of chemicals“, „A mouse. A laser beam. A manipulated memory“, „3 clues to understanding your brain“, „A neural portrait of the human mind“, „Growing evidence of brain plasticity“, „How we read each other's minds“, „I am my connectome“
- [Topics for text "A monkey that controls a robot with its thoughts. No, really."](#)

4 -> and, light, video, now, game, through, show, different, also, look
14 -> and, technology, computer, new, use, machine, now, system, design, build
19 -> now, thing, little, got, something, lot, look, kind, doing, say
25 -> brain, and, neurons, brains, memory, consciousness, autism, cells, mind, sleep

Topics generated from TED Subtitles

0 car, cars, fly, miles, road, drive, around, driving, only, light
1 and, percent, there, than, now, today, two, year, change, look
2 music, sound, play, and, applause, song, now, hear, voice, thank
3 and, africa, india, countries, there, children, country, poor, many, African
4 and, light, video, now, game, through, show, different, also, look
5 and, city, building, cities, new, space, design, built, buildings, project
6 laughter, applause, now, three, two, yeah, four, good, five, okay
7 and, cells, body, stem, cell, tissue, cancer, skin, blood, legs
8 and, feel, life, human, self, social, good, being, happiness, experience
9 she, her, women, and, men, woman, girls, girl, young, sex
10 his, she, him, there, did, went, back, her, after, didn
11 data, internet, information, media, online, web, and, google, new, now
12 and, species, dna, life, human, animals, humans, evolution, bacteria, different
13 and, there, now, than, even, many, question, different, example, much
14 and, technology, computer, new, use, machine, now, system, design, build
15 food, and, eat, bees, plants, plant, feed, farmers, eating, bread
16 and, there, ocean, water, sea, fish, down, ice, animals, species
17 nand, nto, nof, nthat, nthe, nin, nis, nfor, nwith, nwe
18 and, god, his, story, say, man, love, book, him, says
19 now, thing, little, got, something, lot, look, kind, doing, say
20 school, kids, and, children, students, education, child, she, learning, teachers
21 and, energy, water, oil, use, need, carbon, climate, power, gas
22 life, day, even, every, need, thank, applause, many, than, most
23 and, money, business, dollars, companies, company, need, market, work, value
24 and, art, design, work, made, kind, laughter, film, sort, new
25 brain, and, neurons, brains, memory, consciousness, autism, cells, mind, sleep
26 and, power, country, political, china, chinese, democracy, government, american, states
27 war, violence, security, there, police, military, killed, afghanistan, conflict, peace
28 and, universe, there, earth, space, planet, life, stars, mars, sun
29 and, health, cancer, disease, patients, care, medical, patient, drug, treatment

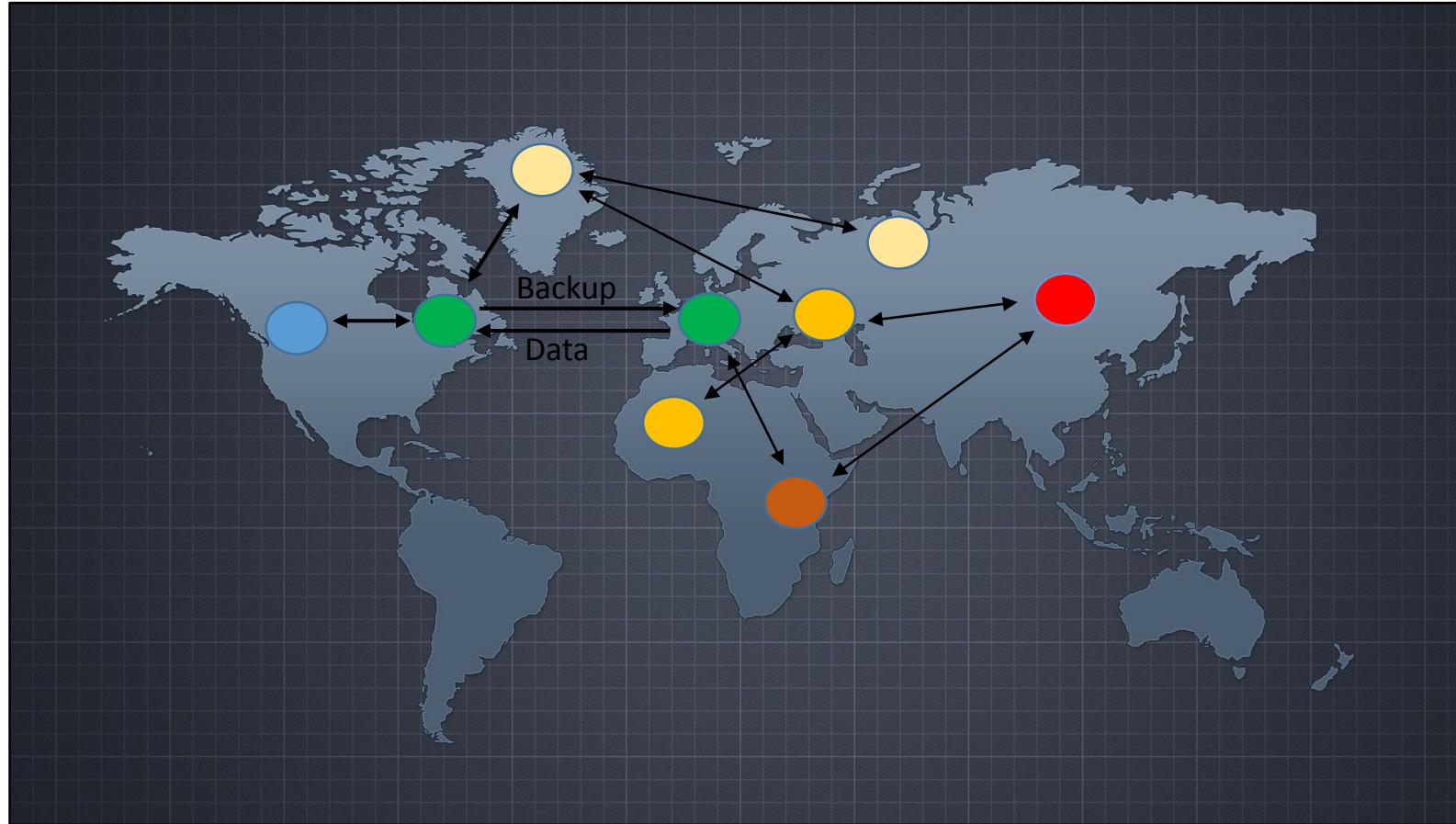
Topics generated from arab TED Subtitles

الأطفال, التعليم, المدرسة, طفل, الطلاب, الطفل, المدارس, مدرسة, تعليم, هؤلاء
 الإنترن特, المعلومات, البيانات, الانترن特, الشبكة, موقع, التكنولوجيا, الكمبيوتر, شبكة, جهاز
 النموذجي, الن, الحمض, الت, سنة, الس, الجينات, الجينوم, البكتيريا, الص
 بأن, الحياة, الأشخاص, لنا, فإن, علينا, أنا, بالنسبة, تكون, مما
 النساء, الرجال, المرأة, الجنس, الفتيات, امرأة, نساء, الجنسية, الزواج, الحب
 والتي, مرة, قمنا, ومن, تقوم, القيام, وهي, استخدام, مختلفة, يقوم
 الصور, صورة, الصورة, القصص, الفيلم, قصة, صور, التصوير, فيلم, الأفلام
 الموسيقى, موسيقى, الصوت, صوت, الروبوتات, الأصوات, الموسيقية, فيديو, موسيقية
 الطاقة, المياه, الكربون, الماء, المحيط, النفط, الغابات, البحر, الحيوانات, طاقة
 عام, المتحدة, أفريقيا, الهند, الولايات, الصين, الدول, سنة, عدد, دولة
 اللغة, كلمة, الكلمات, لغة, الكتب, الإنجليزية, المروor, الكتابة, الصينية
 حسنا, حين, إذن, وبالتالي, كذلك, نعم, بذلك, الجمهور, القيام, ثلاثة
 حسنا, أعتقد, إنها, إنه, أنها, أيضا, أحد, حقا, آخر, لأنه
 السرطان, مرض, المرضي, الخلايا, القلب, العلاج, المرض, الصحية, الرعاية, المريض
 لوحة, متحف, الرسم, amp, quot, التصميم, الفن, تصميم, العمل, نوع
 الحرب, المتحدة, الولايات, العنف, الحكومة, السجن, القوة, الشرطة, السلطة, السلام
 أنني, لدى, على, بدأ, نفسي, حياتي, أستطيع, أنا, وأنا, أريد
 اللعب, الألعاب, اللعبة, لعبة, ألعاب, لعب, كرة, هناك, اللاعبين, ذلك
 الطعام, النحل, الغذاء, النباتات, الحشرات, النمل, الجذع, النباتات, الزراعة
 الدماغ, الخلايا, العصبية, المخ, علم, دماغ, الذاكرة, الأعصاب, خلية, العقل
 الأرض, قدم, المحيط, البحر, تحت, فوق, الفضاء, عبر, الجليد, الماء
 الله, التعاطف, الدين, الدين, الدين, الموت, الحياة, الكتاب, قال, الدينية, والتراث
 دولار, المال, شركة, الشركات, العمل, الأعمال, السوق, التجارية, مليون, الحصول
 المدينة, المدن, مدينة, المبنى, بناء, المباني, السيارات, نيويورك, البناء, المكان
 وكان, يكن, قال, كانوا, الرجل, عام, هؤلاء, منذ, يوم, سنوات
 أوا, إلىn, أن, حسن, شكرn, التيn, علىn, من, أيههn, فيn
 الكون, الفضاء, الأرض, الحياة, النجوم, الفيزياء, نظرية, الشمس, المادة, الكواكب
 التي, علي, الذى, إلى, الأن, لأبد, شيئاً, هذه, حتى, آخر
 اذا, أنها, ان, أنا, اكثـر, ايضاً, الاشياء, اعتقد, حسنا, انهم
 لكي, ومن, بصورة, الامر, والتي, وان, أنها, وهي, لأن, فحسب

Topics generated from engl. Perseus

0 ref, therefore, equal, square, elem, targorder, target, let, straight, angle
1 argument, greek, cit, words, word, used, says, called, marcus, meaning
2 chap, isoc, loeb, must, state, good, things, because, others, each
3 xml, grc, says, blockquote, called, fish, wine, very, having, book
4 things, good, nor, how, can, say, thing, must, like, another
5 xml, letter, salute, opener, rome, dateline, caesar, date, cicero, think
6 pers, ethnic, reg, para, europe, greece, asia, egypt, nation, Africa
7 direct, blockquote, reason, greek, plutarch, moralia, soul, frag, because, earth
8 did, hir, she, thou, shee, bee, thy, mee, thee, uponn
9 caesar, himself, pompey, antony, soldiers, enemy, since, sent, came, army
10 people, could, over, without, while, once, long, very, under, every
11 term, son, thou, she, zeus, odysseus, come, achaeans, did, thee
12 too, chap, known, called, very, mentioned, taken, however, says, probably
13 water, thus, body, can, through, sun, down, each, earth, fire
14 athenians, ships, athens, lacedaemonians, allies, war, para, margin, athenian, sea
15 son, paus, hom, called, para, say, daughter, zeus, scholiast, she
16 para, persname, loeb, surname, cyrus, enemy, army, way, king, come
17 yahweh, israel, king, son, land, house, children, people, god, sons
18 athens, law, loeb, father, money, laws, philip, did, demosthenes, property
19 date, default, king, alexander, battle, since, war, plut, direct, lacedaemonians
20 god, lord, come, don, things, let, like, therefore, jesus, know
21 athenians, both, galleys, having, war, came, themselves, lacedaemonians, went, army
22 enemy, rome, war, senate, consul, camp, army, roman, romans, two
23 lemma, targorder, ditto, root, tln, she, here, house, why, know
24 himself, book, well, two, very, called, love, says, she, used
25 day, patient, food, must, disease, fever, blood, parts, cases, pain
26 marg, date, philip, king, rome, romans, antiochus, war, having, aetolians
27 called, jones, country, stadia, sea, near, date, river, far, situated
28 she, chorus, gods, nor, love, father, thou, thy, yet, where
29 loeb, socrates, greek, xml, grc, plato, date, laws, must, page

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>

Jochen Tiepmar
ScaDS BigData Competence Centre
Ritterstrasse 9-13, 2.0G
04109 Leipzig
jtiepmar@informatik.uni-leipzig.de

This work was funded by the
German Federal Ministry of Education and
Research within the project

Competence Center for Scalable Data Services and Solutions (ScaDS) Dresden/Leipzig
(BMBF 01IS14014B)