

Semantic Web und Ontologien im Bereich des Historical Text Re-use

Semantic Web Applications in the Humanities
Dr. Marco Büchler

Göttingen Centre for Digital Humanities
10. März 2015

Was ist (Historical) *Text Re-use*?

Allgemein: Text Re-use beschreibt die mündliche und schriftliche Wiederverwendung von Textinhalten.

Beispiele: Zitierabhängigkeiten, Paraphrasen aber auch Übersetzung

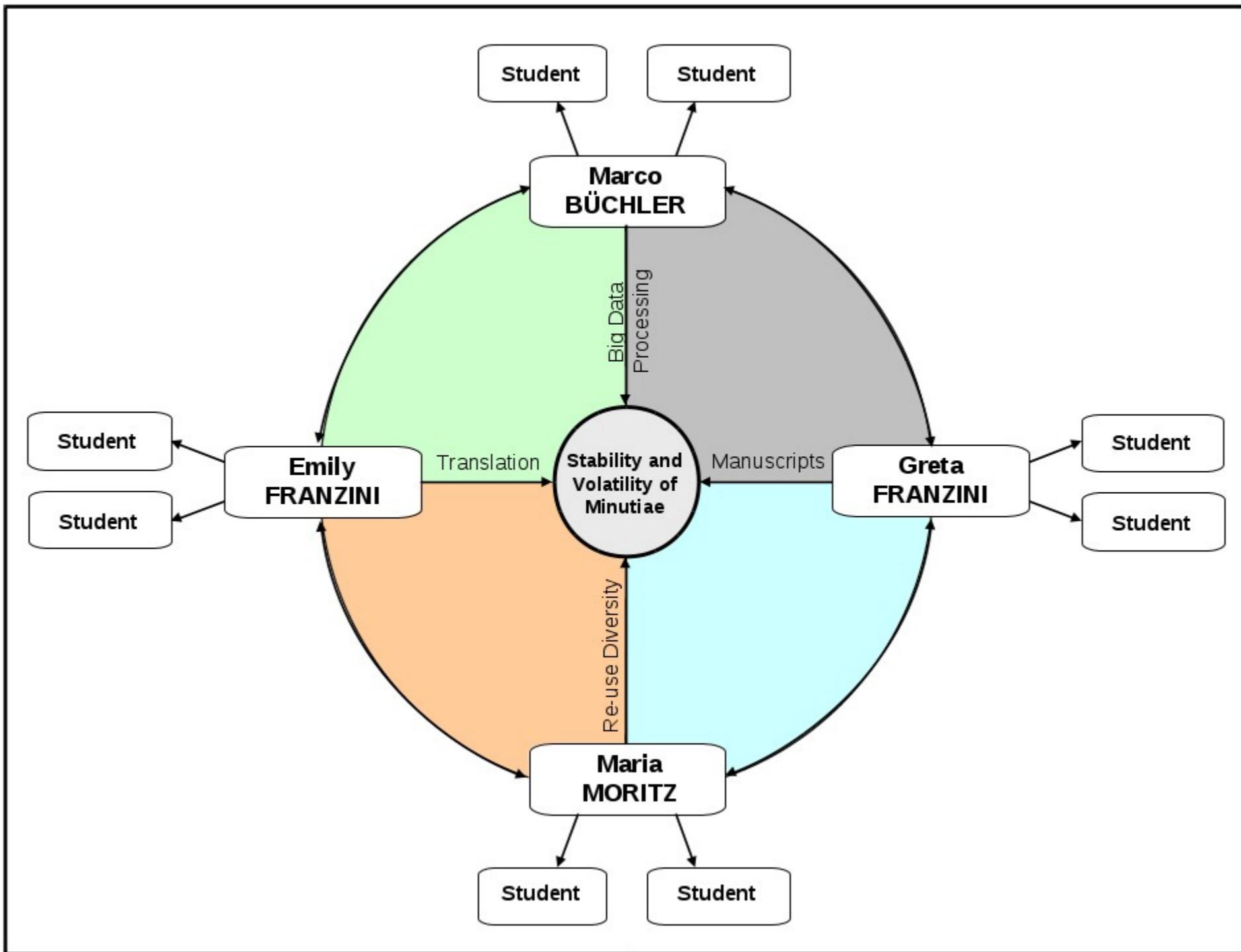
Historische Veränderungen: Sprachevolution, verschiedene Dialekte, “Rechtschreibfehler” bzw. Kopierfehler

Text Re-use in der Welt des Semantic Web



eTRAP

electronic Text Re-use Acquisition Project



Agenda

- A) Zugriff auf **Texte** im Sinne des Semantic Web
- B) Existierende **Knoten- und Kantenontologien**
- C) Unsere **Anforderungen** an Knoten- und Kantentypologien
- D) **Semantic Web** im weiteren Sinne
- E) **Probleme** der automatischen Anreicherung

Zugriff auf Texte und Textteile

Ausgangslage: Wie verweisen wir auf Texte und Textstellen?

Chron. I, 8—10.

11

praeter regionem Edom, quam Esau frater incolebat, iter desti-
naret, missis prius legatis et muneribus suspectum sibi fratrem
explorauit. tum ille obuiam fratri processit: nec tamen Iacob se
ultra fratri credidit. sed pridie, quam inter se fratres conuenirent, 6
5 Deus humana specie assumpta, collectatus cum Iacob refertur.
et cum aduersus Deum praeualuisset, tamen non esse mortalem
non ignorauit: benedici se ab eo flagitabat. tum a Deo ei immu- 7
tatum nomen est, ut ex Iacob Israël diceretur. sed cum ille ui-
cissim a Deo nomen Dei quaereret, non esse querendum, quia
10 admirabile esset, responsum. ex collectatione autem latitudo
feminis Iacob obtorpuit.

Igitur Israël declinans fratris domum, promouit agmen in 10
Salem Sicimorum oppidum, atque ibi loco pretio accepto taberna-
culum statuit sibi. huic oppido Emor, Chorraeus princeps, pree-
15 erat. huius filius Sychem Dinam filiam Iacob ex Lia genitam 2
stupro subdidit. quo comperto Symeon et Leui, Dinae fratres,
omnes in oppido sexus uirilis dolo peremerunt atque impigre so-
roris ulti iniuriam: oppidum a filiis Iacob direptum praedaque
omnis abducta. id factum aegre admodum tulisse Iacob traditur. 3
20 mox a Deo monitus Bethel petiit ibique altarium Deo statuit.
inde in parte turris Gader tabernaculum fixit. Rachel ex parti
obiit: puer Benjamin uocatus est. Isaac decessit annos natus cen-
tum et octoginta. at Esau potens diuitiis erat, uxoribus sibi etiam 4
e Chananaeorum gente assumptis: cuius progeniem in hoc tam
25 praeciso opere inserendam non putaui. si quis studiosior erit, ad
originem reuertatur. post excessum patris Iacob in solo eo, in quo

12 Gen. 33. 16 Gen. 34. 20 Gen. 35. 22 Gen. 36.

3 fratri *Vonckio glossa uidetur* 5 Dominus b || asumpta P 6 Domini-
num b 7 ignorans *Fabricius*; contra de *Prato* malebat ideoque benedici
uel et ben. || se P: sibi b || a Deo b: abeo P || ei P: illi b 9 a Domino
nomen Domini b 10 responsum est b 11 femoris b 14 Chorreus *Si-*
gonius: chorrens P (rrens in rasura, sed prima m. ita scriptum): Chemo-
reus b; LXX in Gen. 34, 2: Ἐμμώρ ὁ Εὐαῖος, ὁ ἀρχῶν τῆς γῆς cf. tamen
de *Prato* ad h. l. 15 Sichem b, at cf. LXX 16 Simeon edd. 17 atq.
nos: at P, ac b 18 ulti sunt b 19 est abducta b 20 Domino b
|| Bethel *Drusius*: Bethlehem Pb; cf. *Genes.* 35 || altare Domino b
23 octaginta P || diuitiis b: diuinitus P 24 chananeorum gentē P
25 miserendam P 26 originem P: origines b

12

Sulpicii Seueri

Isaac uixerat, commorabatur. filii eius aliquando ab eo pascui
gratia cum gregibus secesserunt, Ioseph tamen et Benjamin par-
6 uus domi resederant. carus admodum Ioseph patri obque id
inuisus fratribus, simul quia frequentibus eius somniis maiorem
eum omnibus futurum portendi uidebatur. igitur ad inspiciendos
greges reuisendosque fratres a patre missus oportunus iniuriae
7 fuit. namque uiso fratre consilium necis eius ceperunt. sed
obstante Ruben, cui a tanto facinore abhorrebat animus, in
lacum demissus: mox suadente Iuda deducti ad mitius consilium
negotiatoribus eum, qui tum Aegyptum petebant, uendiderunt. 10
atque ab his Petefrae, preeposito Pharaonis, traditus est.

11 Per idem tempus Iudas, Iacob filius, Sauam Chananaeam
in matrimonium assumpsit. ex qua tres filios sustulit, Her, Onan,
Selam. sed Her Thamar concubio sociatur. quo mortuo Onan fra-
tris uxorem accepit: qui, quia spermata in terram effunderet, ex- 15
2 tinctus a Deo refertur. tum Thamar meretricio habitu assumpto
3 socero mixta est, exque eo geminos edidit. in partu autem illo
mirabile fuit, quod cum prodeundi puer, ad dinoscendum qui
prius nasceretur, obstetrix manum eius coco illigasset, reductus
in aluum matris puer posterior editus est. nomen infantibus 20
4 Fares et Zara inditum. at Ioseph, cum benigne a curatore regio,
qui eum pretio acceperat, haberetur domumque eius et familiam
procuraret, decorus ipse insigni specie uxoris domini oculos in se
conuerterat. cumque amore turpi deperiret, appetitum saepius
nec adquiescentem sibi falso scelere infamat, ac uiro queritur stu- 25
prum sibi intentatum. ita Ioseph in carcerem coniectus. erant in

2 Gen. 37. 12 Gen. 38. 20 Gen. 39. 26 Gen. 40.

1 aliquando nos: aliquanto P, aliquantum b 2 cum] oum P || seces-
serant b || et P: ac b 6 opportunus b 7 cooperant P 11 petefre P
Potifari b; illud reponendum esse docuit de *Prato* coll. LXX atque Hiero-
nymo in Quaest. hebr. et August. in Quaest. in Gen. || preeposito coquo-
rum Phar. coni. de *Prato* cum LXX et Aug. Q. in Gen. 77 12 Sauam
Drusius: suam Pb, cf. LXX et August. Q. 128 in Gen. || chananeam P 13 Ἡρ, Αὐράν, Σηλώμ LXX 14 concubio P: connubio b
15 quia spermata b: qui afer uasta P 16 a Domino b 17 illud
malebat de *Prato*, ac paulo post prodeunte 20 posterior editus *Drusius*:
postero die reductus P, postero die editus b 23 decorūs P 24 turpi P:
ipsius b 26 est coniectus b || erant P ead. m. ex irati corr.

Canonical Text Service (OHC02): Vom Text zur Unique ID

Work identifiers: urn:cts:greekLit:tlg0012.tlg001

Passage citation: urn:cts:greekLit:tlg0012.tlg001:1.10

Range: urn:cts:greekLit:tlg0012.tlg001:1.10-1.20

Words: urn:cts:greekLit:tlg0012.tlg001.mth-01:1.1@Achilles

Words: urn:cts:greekLit:tlg0012.tlg001.mth-01:1.1@Achilles[1]

Word range:

urn:cts:greekLit:tlg0012.tlg001.mth-01:1.1@Achilles-1.10@Atreus

Canonical Text Service Protokoll: Von der Unique ID zur Textpassage

- **GetCapabilities** (listet alle Autoren und Werke mit ihren IDs auf)
- **GetValidReff** (liefert alle validen Referenzen zu einem Werk zurück)
- **GetPassage** (liefert die angeforderte Textpassage zurück)

Canonical Text Service Protokoll: Implementierungen

Implementierung	Kerntechnologie
HomerMultiText (Smith & Blackwell)	RDF, SPARQL
Perseus/Alpheios (Almas & Crane)	XML database
Billion Word Project (Tiepmar)	Relational database model (MySQL)

Existierende Knoten- und Kantenontologien

Citation Typing Ontology (CITO): Überblick

- **Current version:** 2.6.3 (22. Mai 2014)
- ... *CiTO, the Citation Typing Ontology, is an ontology written in **OWL 2 DL** to enable characterization of the **nature or type of citations**, both factually and rhetorically, and to permit these descriptions to be **published on the Web**.*
- ... *The citations characterized may be either **direct and explicit** (as in the reference list of a journal article), **indirect** (e.g. a citation to a more recent paper by the same research group on the same topic), or **implicit** (e.g. as in artistic quotations or parodies, or in cases of plagiarism).*

Citation Typing Ontology (CITO): Kantentypen

Object Properties

[agrees with](#) [cites](#) [cites as authority](#) [cites as data source](#) [cites as evidence](#) [cites as metadata document](#)
[cites as potential solution](#) [cites as recommended reading](#) [cites as related](#) [cites as source document](#)
[cites for information](#) [compiles](#) [confirms](#) [contains assertion from](#) [corrects](#) [credits](#) [critiques](#) [derides](#)
[describes](#) [disagrees with](#) [discusses](#) [disputes](#) [documents](#) [extends](#) [gives background to](#) [gives support to](#)
[has citation characterization](#) [has cited entity](#) [has citing entity](#) [has reply](#) [includes excerpt from](#)
[includes quotation from](#) [is agreed with by](#) [is cited as authority by](#) [is cited as data source by](#)
[is cited as evidence by](#) [is cited as metadata document by](#) [is cited as potential solution by](#)
[is cited as recommended reading](#) [is cited as related by](#) [is cited as source document by](#) [is cited by](#)
[is cited for information by](#) [is compiled by](#) [is confirmed by](#) [is corrected by](#) [is credited by](#) [is critiqued by](#)
[is derided by](#) [is described by](#) [is disagreed with by](#) [is discussed by](#) [is disputed by](#) [is documented by](#)
[is extended by](#) [is parodied by](#) [is plagiarized by](#) [is qualified by](#) [is refuted by](#) [is retracted by](#) [is reviewed by](#)
[is ridiculed by](#) [is speculated on by](#) [is supported by](#) [is updated by](#) [likes](#) [obtains background from](#)
[obtains support from](#) [parodies](#) [plagiarizes](#) [provides assertion for](#) [provides conclusions for](#) [provides data for](#)
[provides excerpt for](#) [provides method for](#) [provides quotation for](#) [qualifies](#) [refutes](#) [replies to](#) [retracts](#)
[reviews](#) [ridicules](#) [shares author institution with](#) [shares authors with](#) [shares funding agency with](#)
[speculates on](#) [supports](#) [updates](#) [uses conclusions from](#) [uses data from](#) [uses method in](#)

Sharing Ancient Wisdoms: Übersicht

- ... *The SAWS ontology is our vocabulary for describing the **types of relationships present between or within our texts**. The ontology relationships themselves can be used to annotate manuscripts with these links.*
- ... *Technically, the SAWS ontology reuses and extends the FRBRoo model of bibliographic and cultural heritage information using the Erlangen OWL implementations.*

Sharing Ancient Wisdoms: Kantentypen

- isVerbatimOf
- isVariantOf
 - isShorterVersionOf
 - isLongerVersionOf
 - IsCloseTranslationOf [from one language into another]
 - isLooseTranslationOf
 - IsCloseRenderingOf [from one dialect or genre into another]
 - isLooseRenderingOf

Unsere Anforderungen an Knoten- und Kantenontologien

Ganzheitliche Sicht auf den *Historical Text Re-use: Re-use Types*

Wisdom Quotation Wit Law Saw Verse Parole

Joke Quip Punch Line Platitude Proverb Rant

Slogan Palindrom Meme Mantra Maxim

Sententiae Motto Loanword Koan Legend

Fact Phraseme Idiom Epigram Definition Edition
Paroimia Gnome Bonmot Battle Cry Cliche

Simile Metaphor Ephithet Abstract Adage

Template Pangram Epitome Anagram Flowery Phrase

Triusm Parable Equation Aphorism Apophtegm

- **Stabilität (gelb):** syntaktisch fest vs. semantisch

- **Absicht (grün)**

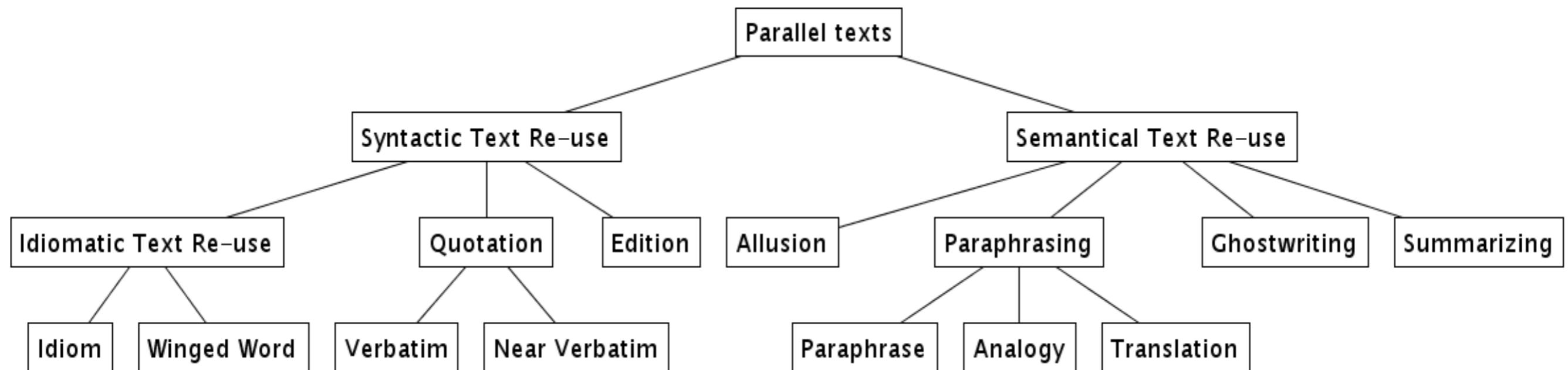
- **Länge des Text Re-use (blau)**

- **Literarische Klassifikation (hell-blau)**

- **Grad der Verbreitung (lila)**

- **Geschriebene vs. gesprochene Sprache**

Ganzheitliche Sicht auf den *Historical Text Re-use: Re-use Styles*



Semantic Web

im weiteren Sinne

Text Re-use Kontext

Source	Target
Aristotle: Odyssey RUID/CTS: ξεῖνε, τὸ μέν σε πρῶτον ἔγων εἰρήσομαι αὐτῇ·τίς πόθεν εἰς ἄνδρῶν; πόθι τοι πόλις ἡδὲ τοκῆες;	Aristotle: Odyssey RUID/CTS: ξεῖνε, τὸ μέν σε πρῶτον ἔγων εἰρήσομαι αὐτῇ·τίς πόθεν εἰς ἄνδρῶν; τίς τοι τάδε εἴματ' ἔδωκεν; οὐ δὴ φῆς ἐπὶ ¹ πόντον ἀλώμενος ἐνθάδ' Ικέσθαι;

Source Target Relation

length idiomacticity expressiousity

Abstract	Adage	Aphorism	Battlecry	Cliche	Commonplace	Ephitet	Epigram
Epitome	Idiom	Kinegram	Loanword	Mantra	Maxim	Metaphor	Motto
Multi word unit	Parable	Parole	Password	Phraseme	Proverb	Quip	Saw
Saying	Slogan	Swearword	Watchword	Winged word	Witticism		

type

alliteration None α β γ δ ε ζ η θ ι κ λ μ ν ξ ο π ρ σ τ υ φ χ ψ ω

metric None

rhyme None

sentiment None + 0 - click related words of source!

audience Scholars

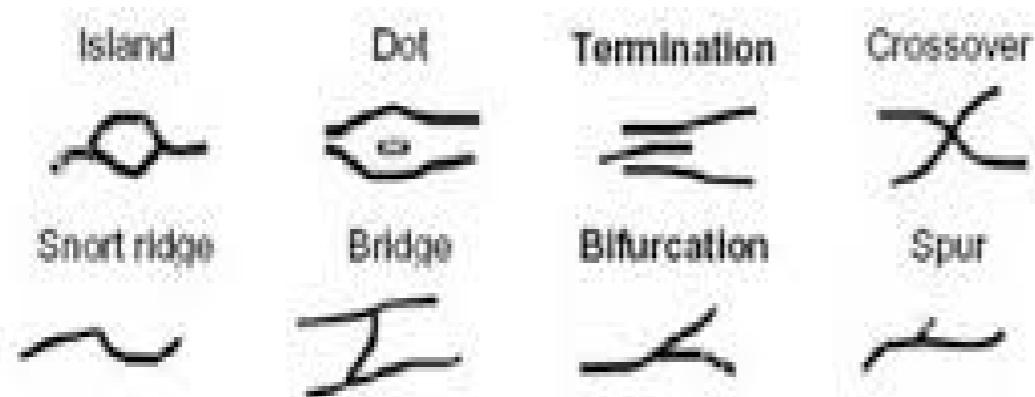
intention Report

archetype

eTRAP: Resulting Questions

Question: What are the common primitives in the re-use diversity?

From biometry (Minutiae):



Problem mit (automatischen) Annotationen

Probleme

- Bestimmung der Richtung bei asymmetrischen Relationen
- Klassifizierung der Knoten- und Kantentypen
- Anforderung der unterschiedlichen kognitiven Tiefe der Knoten- und Kantentypen (aus Sicht der Annotatoren)
- Einbindung von „Fremddaten“ (bspw. Eye-Tracking)
- Umgang mit Unsicherheiten

Linked Data Platform 1.0



Linked Data Platform 1.0

W3C Recommendation 26 February 2015

This version:

<http://www.w3.org/TR/2015/REC-ldp-20150226/>

Latest published version:

<http://www.w3.org/TR/ldp/>

Latest editor's draft:

<http://www.w3.org/2012/ldp/hg/ldp.html>

Test suite:

<https://dvcs.w3.org/hg/ldpwg/raw-file/default/tests/ldp-testsuite.html>

Implementation report:

<https://dvcs.w3.org/hg/ldpwg/raw-file/default/tests/reports/ldp.html>

Previous version:

<http://www.w3.org/TR/2014/PR-ldp-20141216/>

Editors:

[Steve Speicher, IBM Corporation](#)

[John Arwe, IBM Corporation](#)

[Ashok Malhotra, Oracle Corporation](#)

Please check the [errata](#) for any errors or issues reported since publication.

The English version of this specification is the only normative version. Non-normative [translations](#) may also be available.

Copyright © 2015 W3C® ([MIT](#), [ERCIM](#), [Keio](#), [Beihang](#)). W3C [liability](#), [trademark](#) and [document use](#) rules apply.

Abstract

Linked Data Platform (LDP) defines a set of rules for HTTP operations on web resources, some based on RDF, to provide an architecture for read-write Linked Data on the web.

Status of This Document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. A list of current W3C publications and the latest revision of this technical report can be found in the [W3C technical reports index](#) a <http://www.w3.org/TR/>.

This document has been reviewed by W3C Members, by software developers, and by other W3C groups and interested parties, and is endorsed by the Director as a W3C Recommendation. It is a stable document and may be used as reference material or cited from another document. W3C's role in making the Recommendation is to draw attention to the specification and to promote its widespread deployment. This enhances the functionality and interoperability of the Web.

Please see the Working Group's [implementation report](#).

This W3C Recommendation was published by the [Linked Data Platform Working Group](#). Discussions of this document are on public-ldp-comments@w3.org ([subscribe](#), [archives](#)).

This document was produced by a group operating under the [5 February 2004 W3C Patent Policy](#). W3C maintains a [public list of any patent disclosures](#) made in connection with the deliverables of the group; that page also includes instructions for disclosing a

Zusammenfassung

"Stealing from one is plagiarism, stealing from many is research"
Wilson Mitzner, (1876-1933)



SPONSORED BY THE



Visit us via <http://etrap.gcdh.de>