



DIGITAL BREADCRUMBS *of* BROTHERS GRIMM

ENTERING THE THICK FOREST OF INTERCULTURAL TRANSMISSION OF MOTIFS

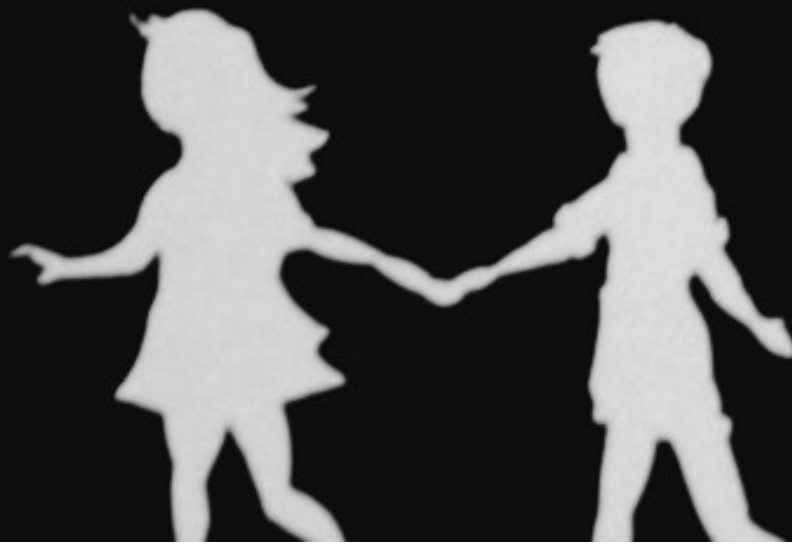
Emily Franzini & Marco Büchler

University of Geneva, January 2016



Electronic Text Reuse Acquisition Project

GÖTTINGEN CENTRE FOR
DIGITAL HUMANITIES



Jacob & Wilhelm Grimm's digital trail

Kinder- und Hausmärchen

1812, 1819, 1837, 1840, 1843, 1850, 1857

Deutsches Wörterbuch

1854

37,000 letters

Marburg Staatsarchiv & Humboldt-Universität zu Berlin



Electronic Text Reuse Acquisition Project
INSTITUTE OF COMPUTER SCIENCE
GÖTTINGEN CENTRE FOR DIGITAL HUMANITIES



7 editions
of the
*Kinder- und
Hausmärchen*

from 1812 to
1857

Why the fairy tales of the Brothers Grimm?

- ✦ Impact on society
- ✦ Big Data
- ✦ Global scope needs an international team
(8 nationalities, 13 languages spoken)
- ✦ Interdisciplinary



Electronic Text Reuse Acquisition Project
INSTITUTE OF COMPUTER SCIENCE
GÖTTINGEN CENTRE FOR DIGITAL HUMANITIES



What are we working on?

Investigating measurable primitives

- ✦ Cultural Studies: tracing **MEMES**
- ✦ Literature: tracing **MOTIFS**
- ✦ Linguistics: tracing **PATTERNS**
- ✦ Computer Science: tracing **MINUTIAE**



Our Projects



DIGITAL BREADCRUMBS *of*
BROTHERS GRIMM



**DATABASE OF
TALE MOTIFS**

**TEXT REUSE
ACROSS GRIMM
TALES
EDITIONS**





DATABASE OF TALE MOTIFS

Motif = “...minimal thematic unit[s]”
(Prince’s Dictionary of Narratology)

Objective = build a database + interface of
the motifs of tales, crossing the language barrier

Why?

- ✦ Investigate&record primitives and their changes
- ✦ Nothing like it exists
- ✦ Humanities: research in folklore
- ✦ Computer Science: algorithmic improvements
 - sharpening the understanding of why and how a text is reused



DATABASE OF TALE MOTIFS

Motif = “...minimal thematic unit[s]”
(Prince’s Dictionary of Narratology)



Snow White





DATABASE OF TALE MOTIFS

Which tales?

- ✦ Snow White
- ✦ Puss in Boots
- ✦ The Fisherman and his Wife

Italian

German

Russian

French

Iranian

English

Rumanian

Spanish



Motif =
“...minimal thematic unit[s]”
(Prince’s Dictionary of Narratology)

Aarne-Thompson (AT) Motif-Index

- A. Mythological Motifs
- B. Animal Motifs
- C. Motifs of Tabu
- D. Magic
- E. the Dead
- F. Marvels
- G. Ogres
- H. Tests
- J. the Wise and the Foolish
- K. Deceptions
- L. Reversals of Fortune
- M. Ordaining the Future
- N. Chance and Fate
- P. Society
- Q. Rewards and Punishments
- R. Captives and Fugitives
- S. Unnatural Cruelty
- T. Sex
- U. the Nature of Life
- V. Religion
- W. Traits of Character
- X. Humor
- Z. Miscellaneous Groups of Motifs



DATABASE OF TALE MOTIFS

Motif =

“...minimal thematic unit[s]”

(Prince's Dictionary of Narratology)

Active Ontology x Entities x Individuals by class x

Class hierarchy Class hierarchy (inferred)

Class hierarchy:

- Thing
 - A. Mythological motifs
 - A0-A99. Creator
 - A0. Creator
 - A1. Identity of creator
 - A1.1. Sun-god as creator
 - A1.2. Grandfather as creator
 - A1.3. Stone-woman as creator
 - A1.4. Brahma as creator
 - A2. Multiple creators
 - A3. Creative mother source of everything
 - A5. Reason for creation
 - A7. Creator descendants
 - A10. Nature of the creator
 - A20. Origin of the creator
 - A30. Creator companions
 - A40. Creator advisers
 - A50. Conflict of good and evil creators
 - A60. Marplot at creation
 - A70. Creator-miscellaneous motifs
 - A100-A599. Gods
 - A1200-A1699. Creation and ordering of human life

Object property hierarchy Data property hierarchy Individuals by type Annotation property hierarchy Datatypes

Object property hierarchy:

topObjectProperty



GRAPH DATABASE (RDF) = VIRTUOSO
QUERY LANGUAGE = SPARQL











DIGITAL BREADCRUMBS *of*
BROTHERS GRIMM
DATABASE OF MOTIFS

Open Search 

Search by:

Title of tale  - 

Motif n.1	Stepmother 
Motif n.2	- 
Author	- 
VIAF number	- 
Language	- 
Date	- 

Results: 7 Results

Tale: Bella Venezia
Author: Calvino, Italo
Collection: Fiabe Italiane
Date: 1956
VIAF: 181208131
Motif: **Stepmother** (P282)
Url: www.text.....

Tale: Schneewittchen
Author: Gebrüder Grimm
Collection: Kinder- und
Hausmärchen
Date: 1819
VIAF: 187449723
Motif: **Stepmother** (P282)
Url: www.text.....

Tale: Schneewittchen
Author: Gebrüder Grimm
Collection: Kinder- und
Hausmärchen
Date: 1837
VIAF: 187449723
Motif: **Stepmother** (P282)
Url: www.text.....



What are we working on?

Investigating measurable primitives

- ✴ Cultural Studies: tracing **MEMES**
- ✴ Literature: tracing **MOTIFS**
- ✴ Linguistics: tracing **PATTERNS**
- ✴ Computer Science: tracing **MINUTIAE**



Typical computer scientists'
expectation: **plagiarism**

TEXT REUSE ACROSS GRIMM TALES EDITIONS



Humanists' expectation: **oversimplification**

**TEXT REUSE
ACROSS GRIMM
TALES
EDITIONS**

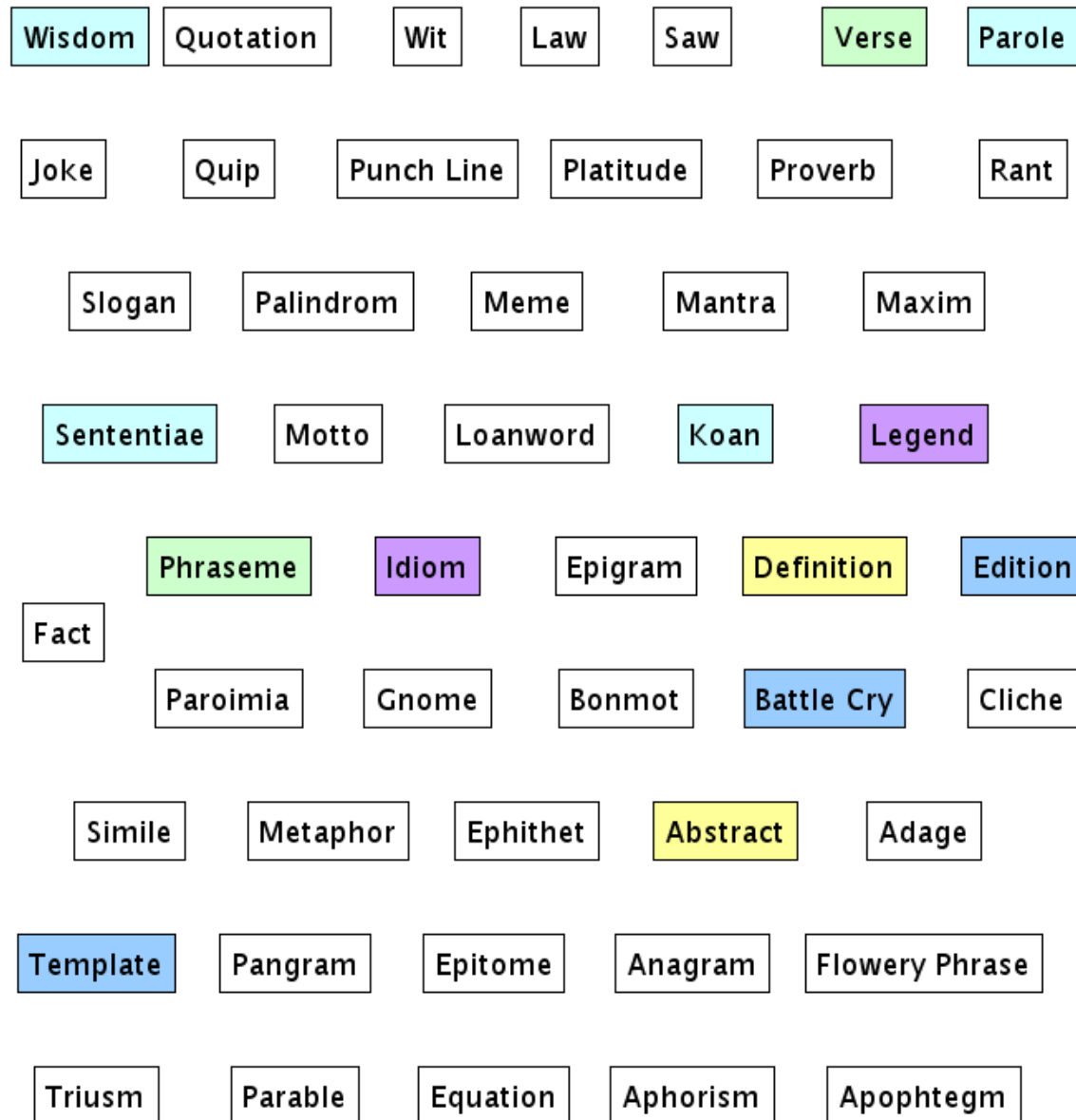




ACID for the Digital Humanities:

-  Acceptance
-  Complexity
-  Interoperability
-  Diversity

ACID for the Digital Humanities – Diversity (Reuse Types)



TEXT REUSE
ACROSS GRIMM
TALES
EDITIONS



Stability

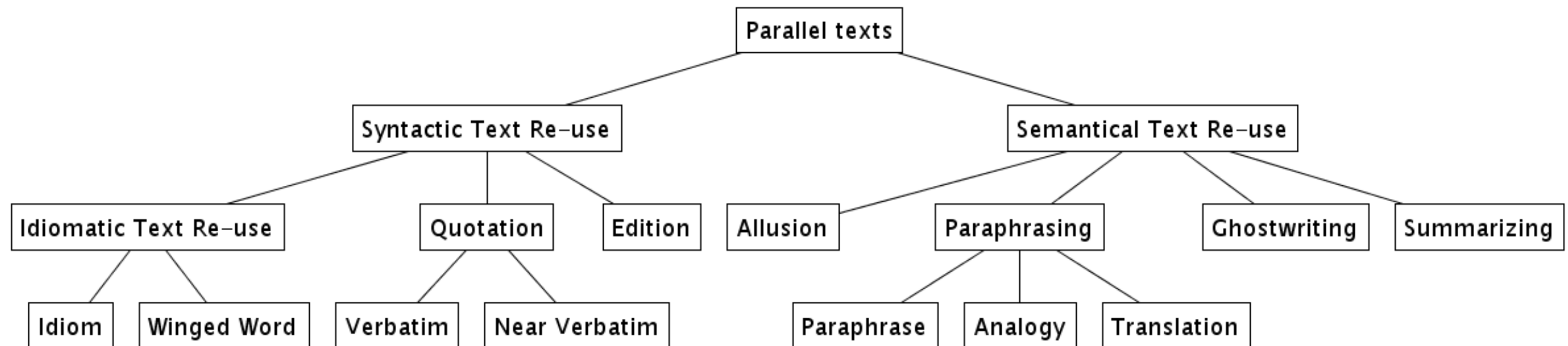
Purpose

Size of text reuse

Classification

Degree of distribution

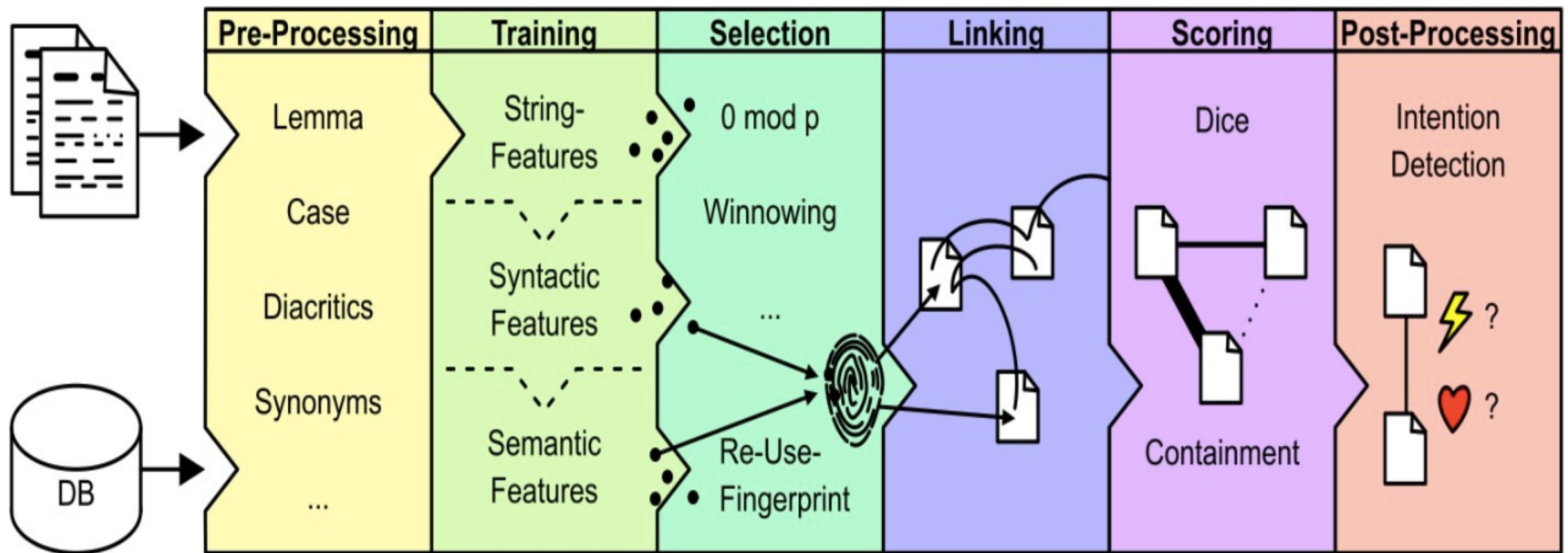
Written and oral transmission



TRACER

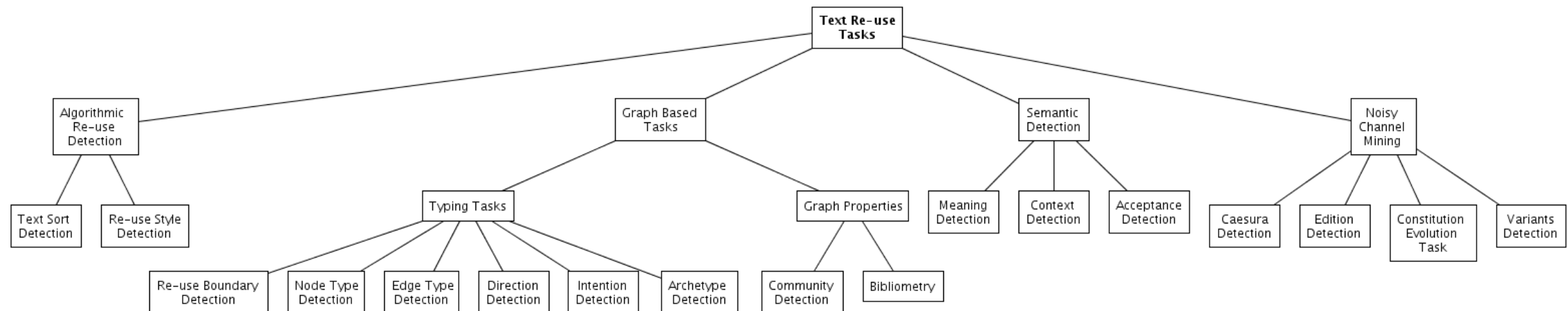
6 steps

TEXT REUSE
ACROSS GRIMM
TALES
EDITIONS



ACID for the Digital Humanities – Complexity

TEXT REUSE ACROSS GRIMM TALES EDITIONS

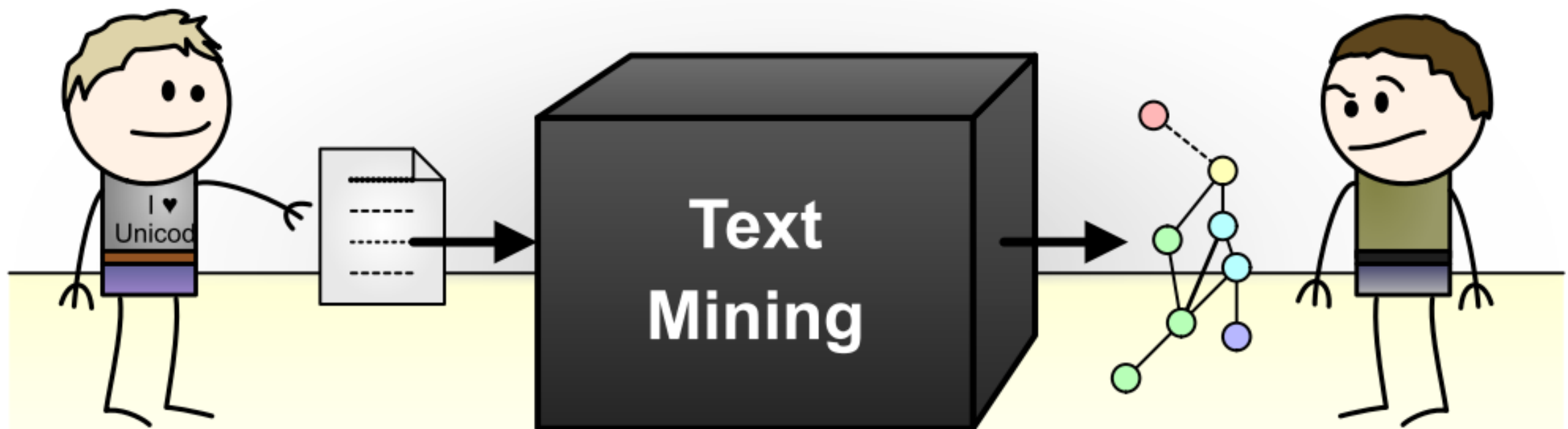




cit-quote-bibl	blockquote	bibl without quote
<pre> <cit> <quote> du/o ku/nes a)rgoi\ ei(/ponto </quote> <bibl n="Hom. Od. 2.11"> Od. 2.11 </bibl> </cit> </pre>	<pre> <quote rend="blockquote"> <line> a)gxou= d' i(stame/nh e)/pea ptero/enta proshu/da <bibl n="Hom. Il. 4.92">Il. 4.92</bibl> </line><line> a)ll' a)/ge nu=n ma/stiga kai\ h(ni/a sigalo/enta <bibl n="Hom. Il. 5.226">Il. 5.226</bibl> </line> </quote> </pre>	<pre> <p> [...]a)nti\ tou= proe/pinon. kuri/ws ga/r e)sti tou=to propi/nein, to\ e(te/rw pro\ e(autou= dou=nai piei=n. kai (*)odusseu\s de\ para\ tw= *(omh/rw <bibl n="Hom. Od. 13.57">Od. 13.57</bibl> [...] </p> </pre>

ACID for the Digital Humanities – Acceptance I

TEXT REUSE
ACROSS GRIMM
TALES
EDITIONS

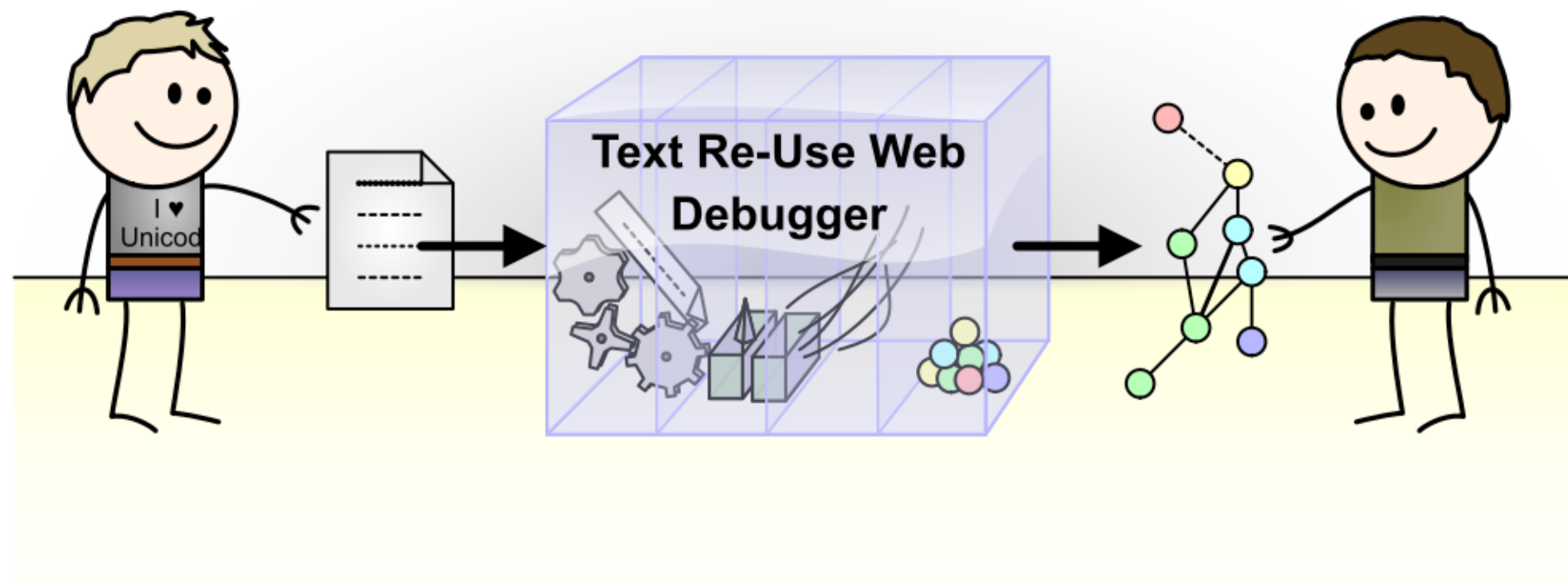




Text Mining



How to be accepted by humanists
if text mining is a black box,
that can't be looked into?



Transparency: How to provide user-friendly insights into complex mining techniques and machine learning?

prev 0 1 2 3 4 5 6 ... 1146 next

ACID for the Digital Humanities – Acceptance V

TEXT REUSE
ACROSS GRIMM
TALES
EDITIONS



Step 0: Searching

Step 1: Preprocessing

Please select a preprocessing strategy:

01:02-WLP:lem=true_syn=false_ssim=false_redwo=false:ngram=5:LLR=true_toLC=true_rDia=false_w2wl=false:wlt=5

change

Unprocessed Sentence:

In the beginning God created the heavens and the earth.

Preprocessed Sentence:

in the begin god create the heaven and the earth .

correct

Your correction for the processed sentence:

in the begin god create the heaven and the earth .

Your comment:

submit changes

Other users preference

No users have suggested a change in the preprocessing level

next Level



ACID for the Digital Humanities – Acceptance VI

TEXT REUSE
ACROSS GRIMM
TALES
EDITIONS



Step 0: Searching

Step 1: Preprocessing

Step 2: Featuring

Please select a training strategy: Bi Gram Shingling Training change

Preprocessed sentence: in the begin god create the heaven and the earth .

Position	Feature
0	in the
1	the begin

next Level

Position	Feature
2	begin god
3	god create

Position	Feature
4	create the
5	the heaven

Position	Feature
6	heaven and
7	and the

Position	Feature
8	the earth
9	earth .



Electronic Text Reuse Acquisition Project
INSTITUTE OF COMPUTER SCIENCE
GÖTTINGEN CENTRE FOR DIGITAL HUMANITIES



ACID for the Digital Humanities – Acceptance VII

TEXT REUSE ACROSS GRIMM TALES EDITIONS



Step 3: Selecting

Please select a selecting strategy: Local Max Feature Frequency Selector:FeatDens=0.8 change

Agenda

- word** = This word belongs to the fingerprint
- word** = This word originally doesn't belong to the fingerprint but was selected by the user to belong to the fingerprint
- word** = This word doesn't belong to the fingerprint
- word** = This word originally belonged to the fingerprint but was selected by the user to not belong to the fingerprint

initial configuration: in the the begin begin god god create create the the heaven heaven and and the the earth earth .
current configuration: in the the begin begin god god create create the the heaven heaven and and the the earth earth .

selected features



not selected features

in the
the begin
god create
the heaven
heaven and
and the
the earth
earth .

begin god
create the

Other users preference

Feature	users selected	users not selected
in the	0	1
the begin	1	0
begin god	1	0
god create	1	0
create the	0	1
the heaven	1	0
heaven and	1	0
and the	0	1
the earth	1	0
earth .	0	1

Statistics

Feature	Selected Features	Total number of features
in the	27114	32227
the begin	470	480
begin god	0	5
god create	27	45
create the	17	38
the heaven	1624	1695
heaven and	389	398
and the	31808	40850
the earth	4776	5222
earth .	1030	1040

next Level

submit changes



Electronic Text Reuse Acquisition Project
INSTITUTE OF COMPUTER SCIENCE
GÖTTINGEN CENTRE FOR DIGITAL HUMANITIES

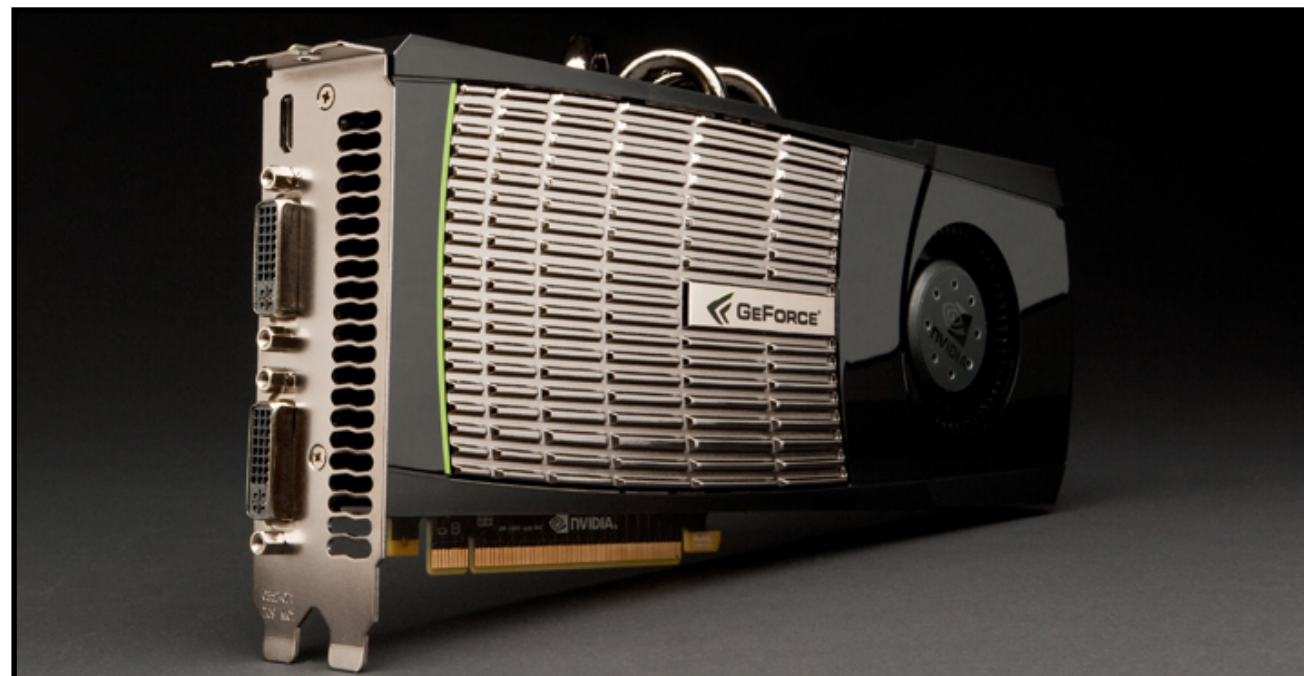


TEXT REUSE ACROSS GRIMM TALES EDITIONS



Further topics

- ✦ Evaluation of text reuse: manual or automatic
- ✦ Research on the measurable primitives including eye-tracking and electroencephalogram
- ✦ Scaling Text Reuse at any size



<http://etrap.gcdh.de/>



*Copying from one is plagiarism,
copying from many....*

*is **research**.*

-Wilson Mitzner-



SPONSORED BY THE

Federal Ministry
of Education
and Research

