



CATMA

An Integrated Environment for Literary Text Annotation and Analysis

Marco Petris, Institute for German Studies, University of Hamburg

Summer SOC, Crete, 18 June 2019

A bit of history

- TACT (Bradley, Lancashire)
- 2008
- 2009 CATMA 1, single user desktop installable for all major OS
- 2011 Google Digital Humanities Award
- 2012 CATMA 4, multi user web application
- 2019 September CATMA 6, versioning and visualization
- 2017-2020 forTEXT
<http://fortext.net>

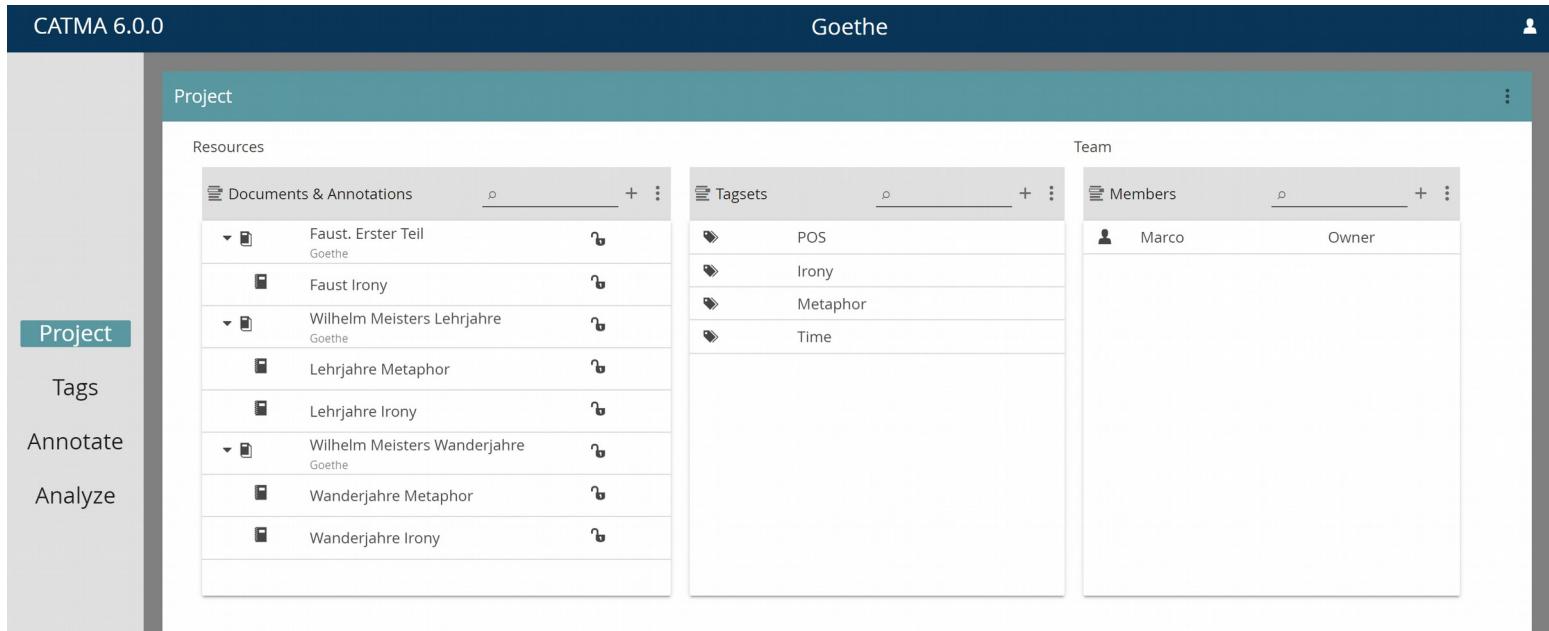


The screenshot shows a command-line interface window titled "Eingabeaufforderung - forcedos usebase.exe D:\Texte\in\PloetzTC.tdb". The window has tabs for File, Select, Displays, Group, and Help (F1). The status bar indicates "191 K o". A "Query Dialog Box" is open in the center, containing the text "Query: .*in; when partofspeech NN". Below this, a large black text area displays XML-like code with entities like > <6>, <RAW rosenhain><POS NN>, and <RAW drin>. The bottom of the window has the instruction "To choose option from Menu Bar, press the SPACEBAR".

```
> <6>    ARTDEF>der : <RAW rosenhain><POS NN>>rosenhain ! <RAW drin><POS
> <11>      PIAT>kein : <RAW steierin><POS NN>>steierin. ! "<RAW
> <13>          : <RAW hohepriesterin><POS NN>>hohepriesterin", <RAW
> <14>      NE>balduin, <RAW cousin><POS NN>>cousin ! <RAW einer><POS
> <16>    ARTIND>ein : <RAW tschechin><POS NN>>tschechin, <RAW denn><POS
```

A CATMA Project

- Documents • Tagsets
- Collections • Team



The screenshot shows the CATMA 6.0.0 software interface with a project titled "Goethe". The left sidebar has tabs for "Project" (which is active), "Tags", "Annotate", and "Analyze". The main area is divided into three sections: "Resources", "Tagsets", and "Team".

- Resources:**
 - Documents & Annotations
 - Faust. Erster Teil (Goethe)
 - Faust Irony
 - Wilhelm Meisters Lehrjahre (Goethe)
 - Lehrjahre Metaphor
 - Lehrjahre Irony
 - Wilhelm Meisters Wanderjahre (Goethe)
 - Wanderjahre Metaphor
 - Wanderjahre Irony
- Tagsets:**
 - POS
 - Irony
 - Metaphor
 - Time
- Team:**
 - Members
 - Marco (Owner)

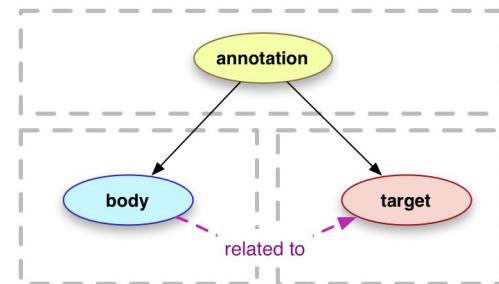
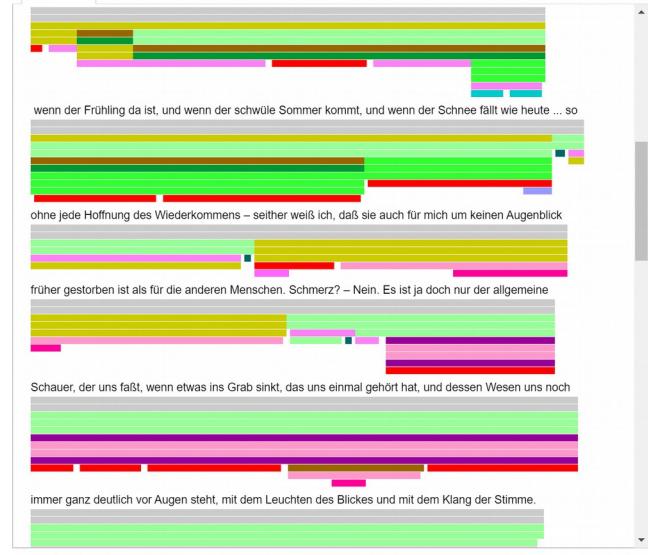
Tags and Tagsets

- Tag
 - name
 - color
 - single rooted tree
 - type-instance-relationship
 - User Defined Properties
- reusable across Projects
- Taxonomy
- multi purpose

Tagsets	Tags	Properties	Values
Time	▼ <ul style="list-style-type: none">▼ timerelation_discours--histoire▼ frequency▼ singulative▼ multi_singulative		
		narrated_content	
		► repetition_count - 1,2...	
		► completeness - complete,incomplete	
	repetitive	► narrated_content,repetition_count	
	iterative		
	▼ duration		
	▼ anisochrony		

Annotations and Collections

- Annotation: set of key/values pairs
 - structural markup
 - hermeneutic or semantic markup
 - comment
- Web Annotation Data Model
 - <https://www.w3.org/TR/annotation-model/>
- Collection as a layer of Annotations
- Document dependent
- each Document can have N Collections
- structure Annotations





The Modules

- Tags
- Annotate
- Analyze



Annotate

CATMA 6.0.0 Goethe

WILHELM MEISTERS WANDERJAHRE x

Project Tags **Annotate** Analyze

ganz nahe liegen. Und was mich noch mehr in meinen höhern Aussichten bestärkte, war jenes Gemälde, das leider nummehr fast ganz erloschen ist. Sobald Sie wissen, was es vorstellen soll, so werden Sie sich's entziffern können, wenn ich Sie nachher davor führe. Dem heiligen Joseph war nichts Geringeres aufgetragen, als ein
Zwischen zwei gegebenen Säulen soll der Prachtsessel
sorgfältig das Maß von Breite und Höhe und arbeitet einen kostlichen Königsthron. Aber wie
erstaunt ist er, wie verlegen, als er den Prachtsessel herbeischafft: er findet sich zu hoch
und nicht breit genug. Mit König Herodes war, wie bekannt, nicht zu spaßen; der fromme Zimmermeister ist in der größten Verlegenheit. Das Christkind, gewohnt ihn überallhin zu
begleiten, ihm in kindlich demütigem Spiel die Werkzeuge nachzutragen, bemerkte seine Noth
und ist gleich mit Rath und That zur Hand. Das Wunderkind verlangt vom Pflegevater: er solle
den Thron an der einen Seite fassen; es greift in die andere Seite des Schnitzwerks und
beide fangen an zu ziehen. Sehr leicht und bequem, als wär' er von Leder, zieht sich der

textfunktionale/illokutionäre Ironie
rhetorische Ironie
romantische Ironie

Collection currently being edited
Wanderjahre JH

direkte Signalisierung
gesellschaftlicher Handlungsbereich
ironisch durch Illokutionsstruktur
thematische Einstellung

Properties

Selected Annotations

Annotation	Tag	Author	Coll	Tagset
Zwischen zwei g...l...t des König	Narrative Ironie	cm	Wand	rhetorische Ironie
Das Christkind,[...]rt und Stelle	Persiflage	cm	Wand	rhetorische Ironie
Zwischen zwei g...l...t des König	Figur	cm	Wand	rhetorische Ironie
Das Christkind,[...]rt und Stelle	Untertreibung/Übertr...	cm	Wand	rhetorische Ironie

?

14 /477

ANALYZE

1 100



Analyze

- Query Language
 - "Mephistopheles"
 - wild="M%"
 - tag="analepse"
 - tag="anachrony/%"
- Query Builder
- Results grouped and counted by phrase or Tag

1. How do you want to search? 2. How does your phrase look like?

The first word starts with <input type="text" value="M"/>	The next word starts with <input type="text"/>
The first word contains <input type="text"/>	The next word contains <input type="text"/>
The first word ends with <input type="text"/>	The next word ends with <input type="text"/>
The first word is exactly <input type="text"/>	The next word is exactly <input type="text"/>
The position of this word is <input type="text" value="1 word after the prev"/> <input type="button" value="Remove"/>	

If your phrase contains more words you can

Show in preview with a maximum total frequency of

Your search

will match for example:

Phrase	Frequency
► Marterort	1
► Marchen	1
► Magister	3
► ...	4



Analyze

Query

freq>0

Execute Query

+ New Query

Query Builder

Wordlist



Result by Phrase

Result by Tag

Phrase	Frequency	Visible in Kwic
► Und	410	<input type="checkbox"/>
► ?	409	<input type="checkbox"/>
► nicht	394	<input type="checkbox"/>
► zu	374	<input type="checkbox"/>
► ist	312	<input type="checkbox"/>
► ;	309	<input type="checkbox"/>
► ein	294	<input type="checkbox"/>
► Mephistopheles	287	<input checked="" type="checkbox"/>
► Faust	265	<input type="checkbox"/>
► das	260	<input type="checkbox"/>
► in	247	<input type="checkbox"/>
Total count: 6.802	Total frequency: 37.916	

Vega Select all Deselect all

Select all for Kwic Deselect all for Kwic

Documents and annotations constraining this search

► Faust. Erster Teil

Document/Annotations	Left Context	Keyword	Right Context	Start
Faust. Erster Teil	oben. Ist gerettet!	Mephistopheles	zu Faust. Her zu	1
Faust. Erster Teil	Mir graut's vor dir.	Mephistopheles	. Sie ist gerichtet!	1
Faust. Erster Teil	habe ich mich übergeben!	Mephistopheles	zu Faust. Komme!	1
Faust. Erster Teil	wäre ich nie geboren!	Mephistopheles	erscheint draußen. Auf!	1
Faust. Erster Teil	Sie streuen und weihen.	Mephistopheles	. Vorbei! Vorbei!	1
Faust. Erster Teil	sich, beugen sich.	Mephistopheles	. Eine Hexenzunft. Faust	1
Faust. Erster Teil	dort um den Rabenstein?	Mephistopheles	. Weiß nicht was sie	1
Faust. Erster Teil	offen Feld . Faust.	Mephistopheles	auf schwarzen Pferden daher brausend	1
Faust. Erster Teil	, und befreie sie!	Mephistopheles	. Ich führe dich und	1
Faust. Erster Teil	Sie soll frei sein!	Mephistopheles	. Und die Gefahr der	1
Faust. Erster Teil	Faust blickt wild umher	Mephistopheles	Greifst du nach dem	1

1 30
5 token(s) context

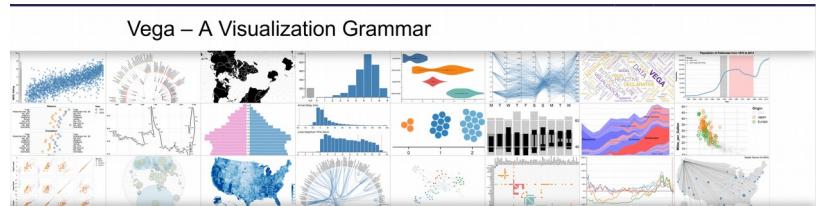
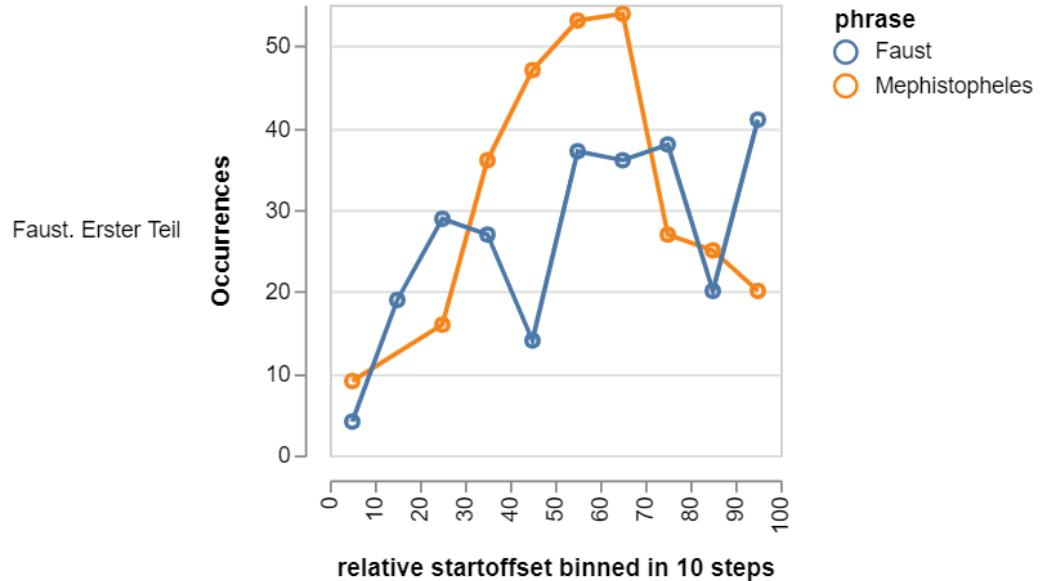
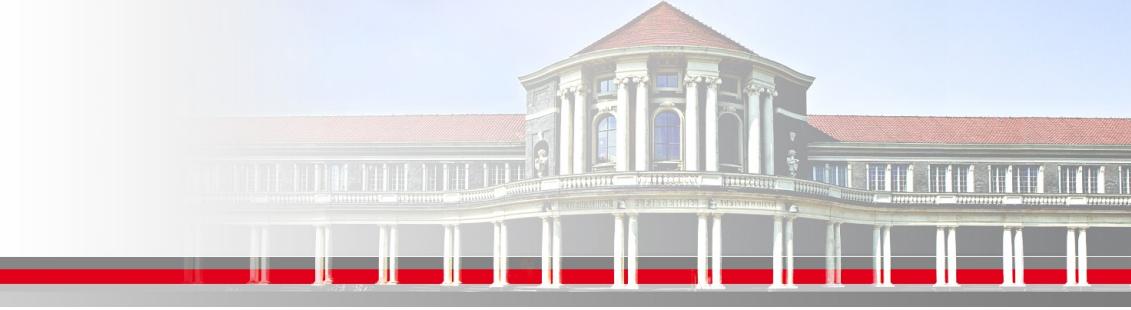
Annotate selected results

Select all



Visualizations

- Plug 'n' Play visualizations
 - Abstract view on Tags and Annotations
 - Flexibility
 - Hidden semantics
- How can visualizations take advantage of the hidden semantics?
- Analogon to the CATMA Query Language: Vega



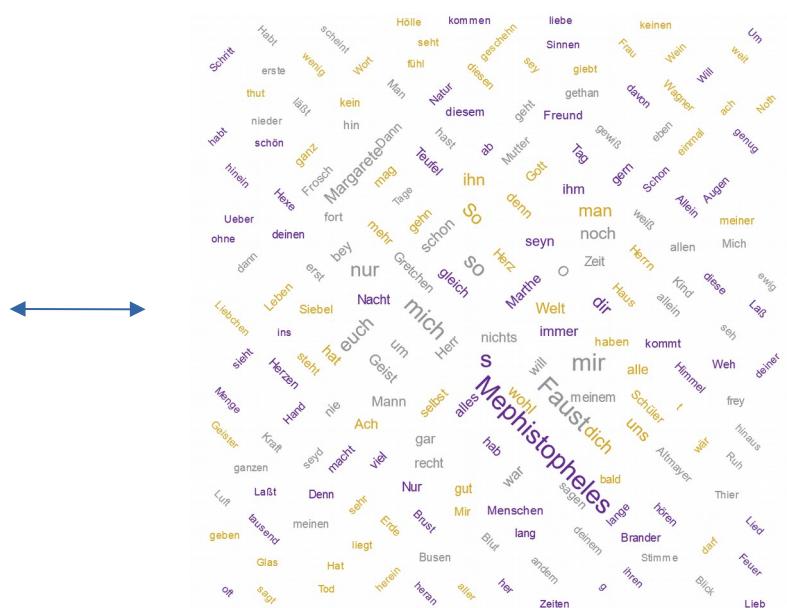
<https://vega.github.io/vega/>



Visualization

- Analogon to the CATMA Query Language: Vega's Visualization Grammar
 - Idea: Analogon to the CATMA Query Builder: a Vega Visualization Builder
 - Idea: Establish visualizations as a fourth resource, tied to their Tagsets

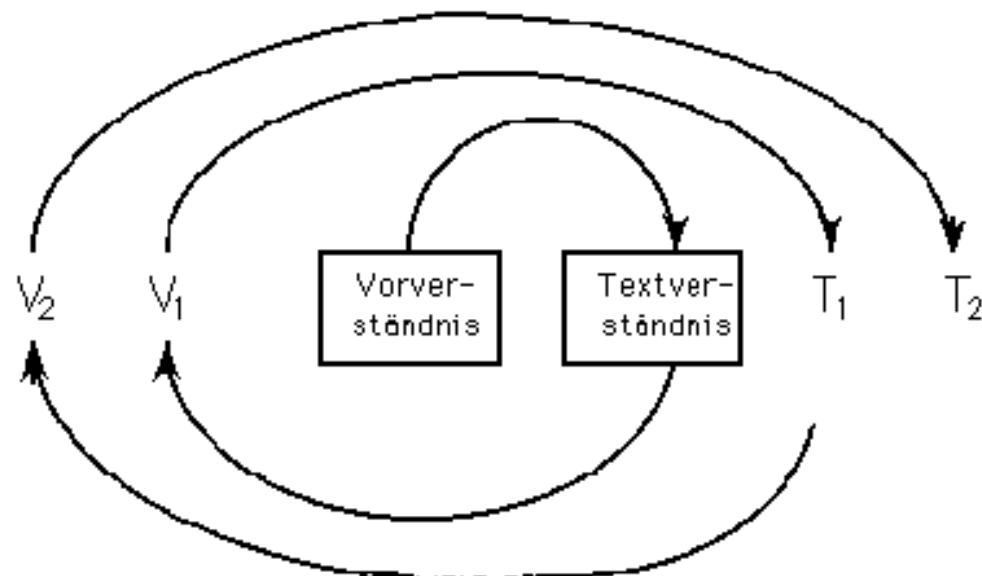
```
59 "enter": {
60   "text": {
61     "field": "phrase"
62   },
63   "align": {
64     "value": "center"
65   },
66   "baseline": {
67     "value": "alphabetic"
68   },
69   "fill": {
70     "scale": "color",
71     "field": "phrase"
72   }
73 },
74 "update": {
75   "fillOpacity": {
76     "value": 1
77   }
78 },
79 "hover": {
80   "fillOpacity": {
81     "value": 0.5
82   }
83 }
84 },
85 "transform": [
86   {
87     "type": "wordcloud",
88     "size": [
89       1000,
90       1000
91     ],
92     "text": {
93       "field": "phrase"
94     },
95     "rotate": {
96       "field": "datum.angle"
97     },
98   }
99 ]
```



[Export as PNG](#) [View Source](#) [Open in Vega Editor](#)



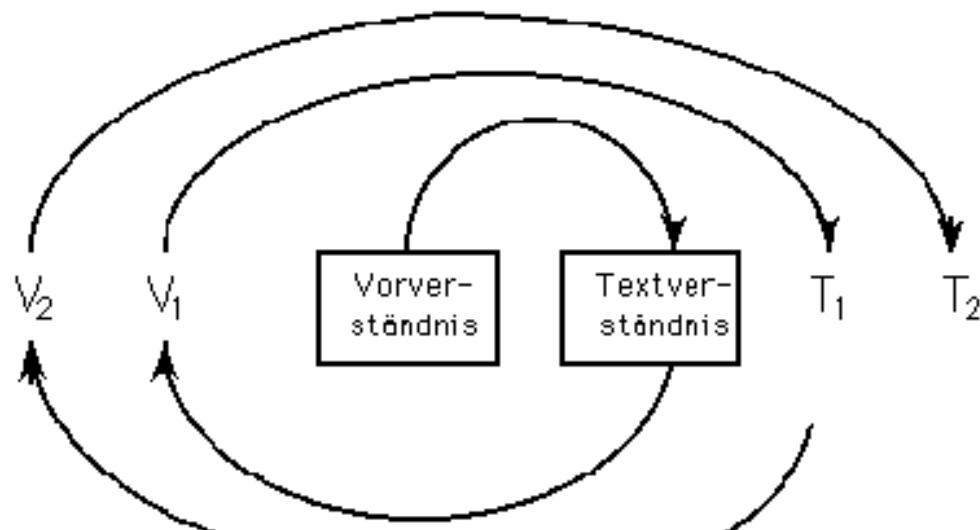
The Workflow



Hermeneutic Circle (Jeßing 2012, 283)



The Workflow



Hermeneutic Circle (Jeßing 2012, 283)

1. Collaborative Environment:
productive use of heterogeneous object descriptions and interpretational approaches

2. Hermeneutic Circle in CATMA:
seamless switching between exploration and annotation of texts

Exploration and Analysis
automated analytic functions
powerful query language
data visualization

3. Fruitful Results:
diverse, well-secured interpretations

Markup and Annotation
freely designable tagsets
non-deterministic standoff markup



Technology

- Gitlab backend
- CATMA Project -> Gitlab Group
- Documents, Collections, Tagsets -> Gitlab Projects
- Project Configuration -> Container Project using git submodules
- Gitlab's User, Role and Permission system
- Vaadin based Frontend using Gitlab's REST API
- JSON, Web Annotation Data Model JSON-LD
- Tinkergraph



"CATMA is the opposite of dogma", urbandictionary.com

Thank you for your attention!

<http://www.catma.de>

<http://www.fortext.de>