

# Development of a web application to manage and edit semantically annotated texts

Thomas Grass, 04. May 2015

Software Engineering for Business Information Systems (sebis)  
Department of Informatics  
Technische Universität München, Germany

[www.matthes.in.tum.de](http://www.matthes.in.tum.de)

master's thesis / information systems

Development of a web application to manage and edit semantically annotated texts

masterarbeit / wirtschaftsinformatik

Entwicklung einer Web-Anwendung zur Verwaltung und Bearbeitung von semantisch annotierten Textsammlungen

project

LexAlyze – Analysis of Legal Texts

student

Thomas Grass

advisor

Bernhard Waltl

supervisor

Prof. Dr. Florian Matthes

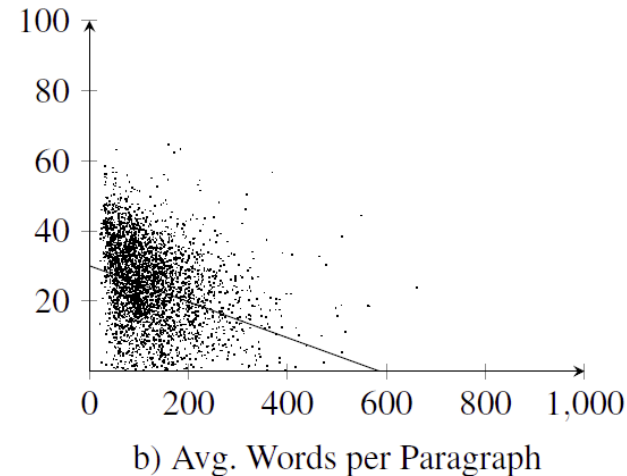
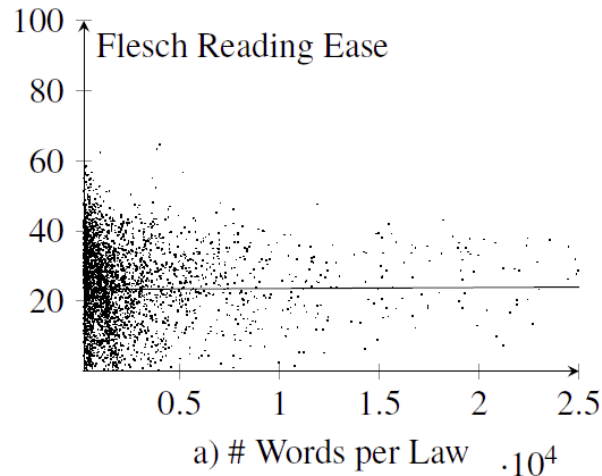
start date

15.03.2015

submission date

15.09.2015

1. Problem / Basic Theory
  1. LexAlyze
  2. Semantically annotated texts
  3. Legal Text Information and Analysis Platform
2. Master's Thesis' Scope
3. Research Questions & Challenges
4. Timeline



## theory

- Legal Texts (e.g. Laws, Judgements) are getting more and more complex
- Usage of quantitative methods of structural network analysis and linguistics provide the possibility to do high class text analysis and comparison of different legal texts
- Project started 2014 by Prof. Dr. Matthes, Prof. Dr. H.C. Grigoleit B. Walzl, and K. Heßler
- Research areas: Law & Software Engineering

## 1. separating texts

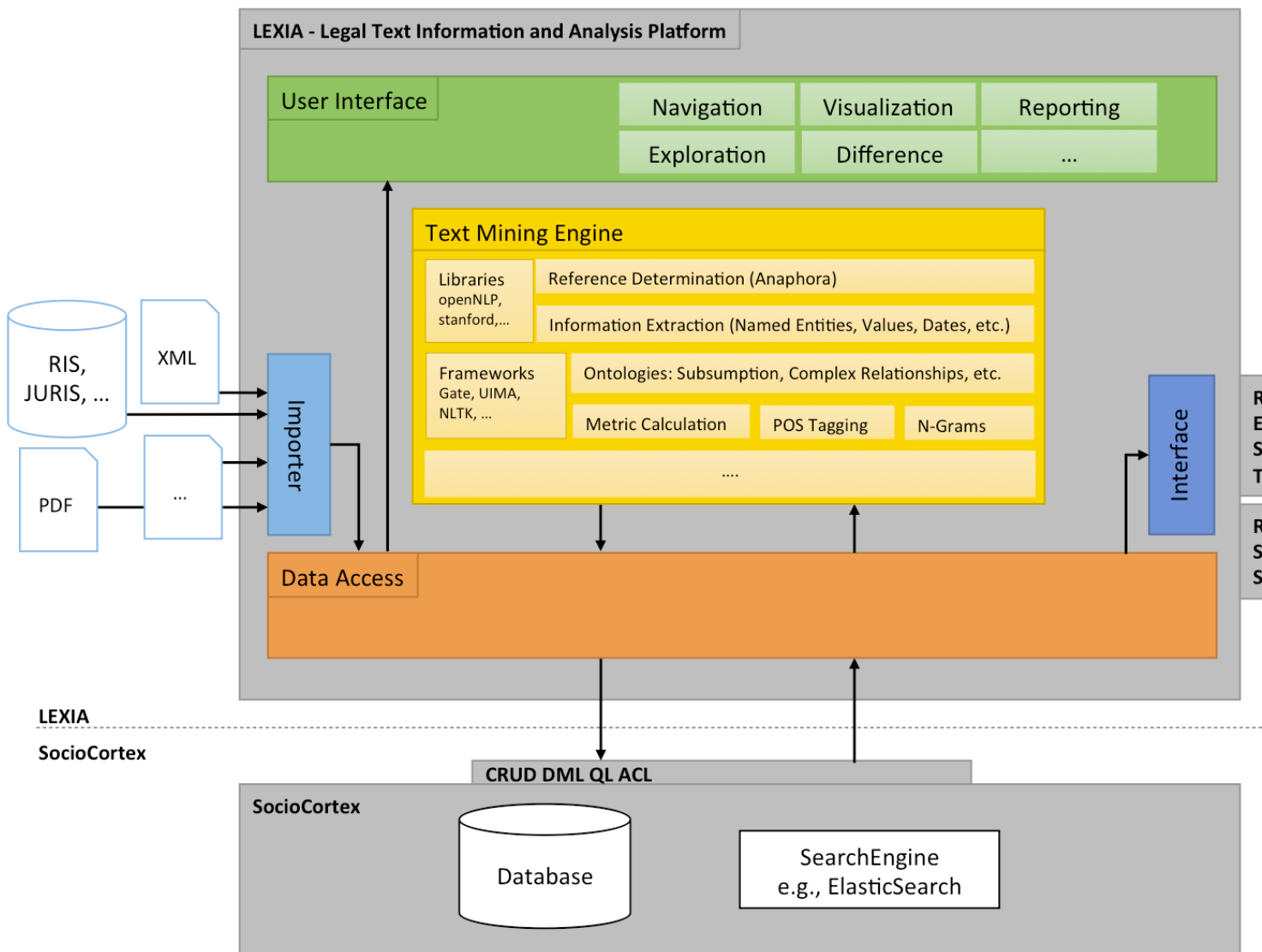
- Texts => Articles
  - Article => Paragraphs
  - Paragraph => Sentences
  - Sentence => Sentence Parts
  - Sentence Parts => Tokens
- „*Grimm’s Fairy Tales*“
  - „*The Frog King*“, „*Rapunzel*“, etc.
  - „*1. In olden times, when wishing still [...]*“
  - „*In olden times,*“, „*when wishing still [...]*“
  - „*In*“, „*olden*“, „*times*“

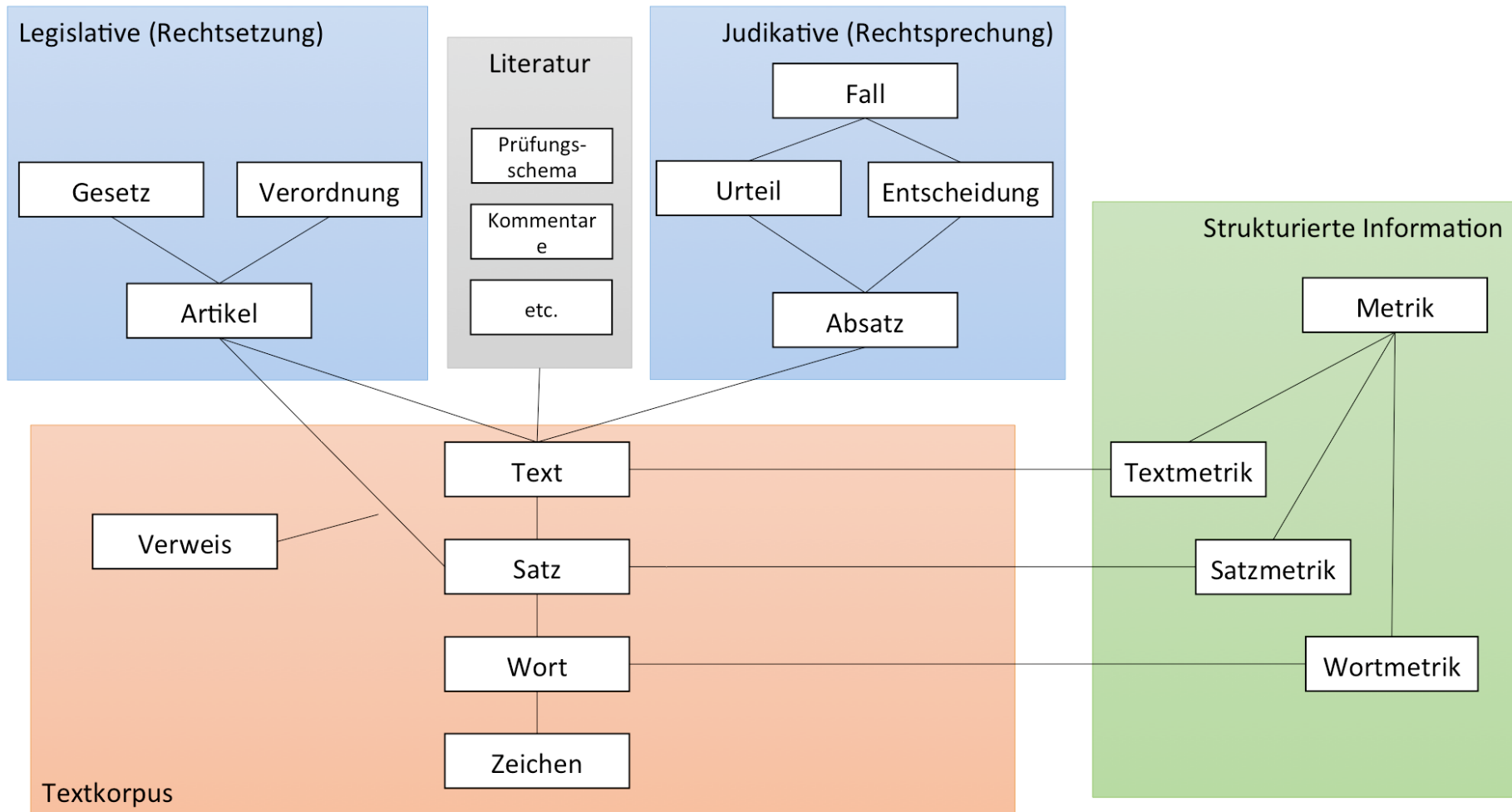
## 2. annotate texts

- e.g. „Homer Simpson likes blue pants.“

```
<sentence>
  <sentence part>
    <subject>
      <noun>Homer</noun><noun>Simpson</noun>
    </subject>
    <verb>likes</verb>
    <object>
      <adjective>blue</adjective><noun>pants</noun>
    </object>
  </sentence part>
</sentence part>
```

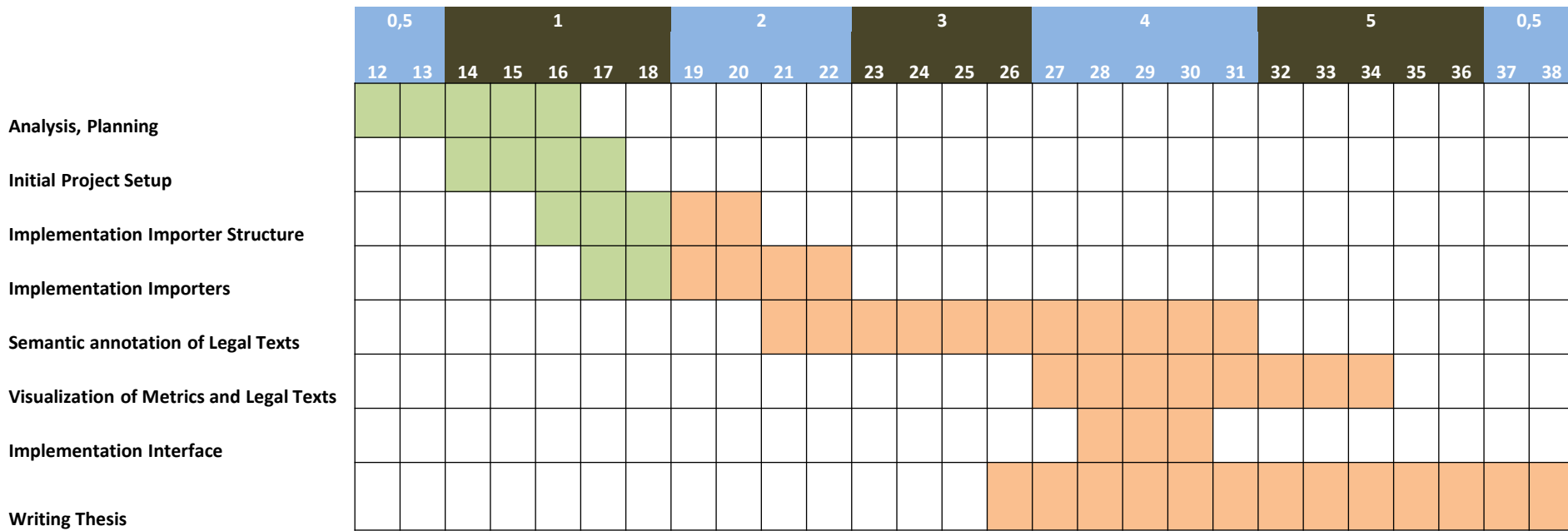
- Web application
- Importing of Legal Texts (e.g. laws, judgements, etc.)
  - Manually by Users
  - Automatically (Push/Pull)
- Extensive text analysis
  - E.g. readability, patterns, etc.
- Persistence of raw texts and analysed annotations, attributes and KPIs
  - SocioCortex
- Visualization of metrics
- Interfaces to provide contents and values for other applications







1. What kind of Legal Texts exist in the German Legislation?
2. What is a way to implement a generic importer for Legal Texts that can easily be adapted?
3. What kind of linguistic text annotations exist and what are benefits and drawbacks of those?
4. How to persist linguistic text annotations in order to access them for further semantic processing?



Thank you for your attention!



**Thomas Grass**  
B.Sc.



Technische Universität München  
Department of Informatics  
Chair of Software Engineering for  
Business Information Systems

Boltzmannstraße 3  
85748 Garching bei München

Tel +49.89.289.17124  
Fax +49.89.289.17136

[thomas.grass@in.tum.de](mailto:thomas.grass@in.tum.de)  
[wwwmatthes.in.tum.de](http://wwwmatthes.in.tum.de)