

Topic Classification for Clauses in Terms of Services with Machine Learning

Bachelor's Thesis Kick-off – Jan Robin Geibel – 15.06.2020

Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
www.matthes.in.tum.de

Outline



1. Motivation
2. Approach and First Steps
3. Outline of the Thesis

- **Terms of Service agreements:** A terms of service agreement **covers a range of issues related to the behavior of a website or service user. E.g., items related to third-party websites, content ownership, copyright notices, payments, additional information [1].**
- **A growing number of Terms of Service agreements** are entered into as more and **more goods and services are being bought online.**
 - E.g., 86% percent of Germans used the Internet in 2019 to purchase goods or services online [2].
- However, studies indicate that most **consumers do not read Terms of Services** [e.g., 3, 4].
 - ➔ The majority of consumers conduct numerous transactions on a daily basis that are governed by Terms of Services without knowing their contents.

Motivation | Sebis-Projects in Regard to Terms of Services

AGB-Check – AI-Supported Legal Review of terms and Conditions to Strengthen Consumer Protection with the goal of:

- developing technologies to support the representation of consumer interests [5].

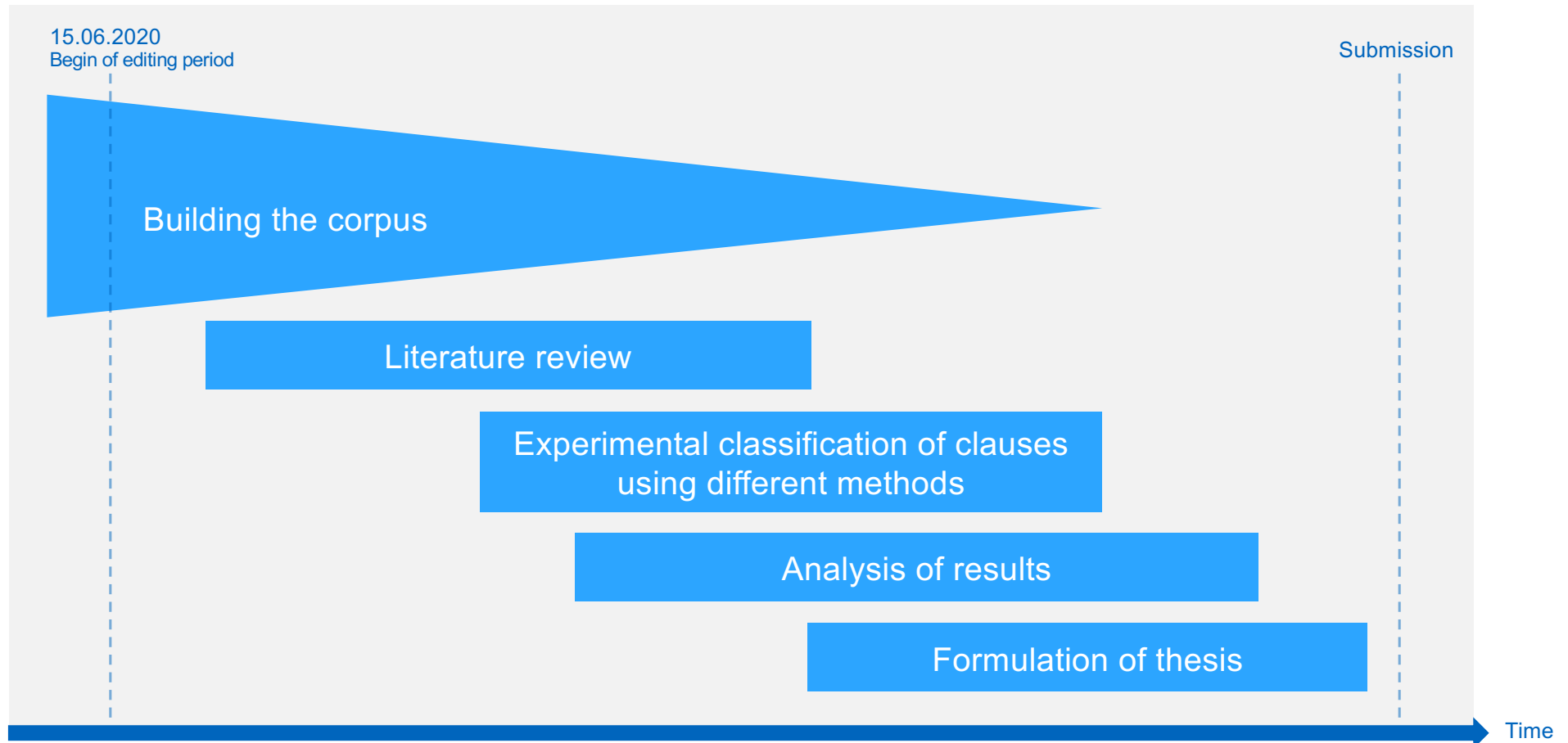
Software Aided Analysis of Terms of Services with the goal of:

- automatically identifying and summarizing Terms of Services
- and assessing them in relation to their lawfulness, customer friendliness and language [6].

Topic Classification for Clauses in Terms of Services with Machine Learning

Exploration of different Machine Learning methods to **automatically identify the topic** being addressed by **individual clauses of Terms of Services**

Approach and First Steps | Overview



The Data

- **Terms of Service agreements from e-commerce websites** provided by the Sebis-Chair:
 - **4875 agreements** in German (html format)
 - **545 agreements** in English (html format)
- A **subset** of those will be **used to construct the corpus**

Approach and First Steps | Building the Corpus

Classes the data is being divided into:

| Clause Label 1 | Clause Label 2 |
|-----------------------|----------------|
| Party | -/- |
| Realization | -/- |
| Contract language | -/- |
| Choice of law | -/- |
| Place of fulfillment | -/- |
| Place of jurisdiction | -/- |
| Dispute resolution | -/- |
| Salvatorius | -/- |
| Prices | -/- |
| Right to refuse | -/- |
| Warranty | -/- |
| Definition | -/- |
| Withdrawal | Right |
| Withdrawal | Consequences |
| Withdrawal | Exclusion |
| Withdrawal | Form |

| Clause Label 1 | Clause Label 2 |
|----------------|----------------|
| Delivery | Liability |
| Delivery | Costs |
| Delivery | Inspection |
| Payment | Time |
| Payment | Default |
| Cancelation | Form |
| Cancelation | Period |
| ... | ... |
| ... | ... |
| ... | ... |
| ... | ... |
| ... | ... |
| ... | ... |
| ... | ... |
| ... | ... |
| ... | ... |
| ... | ... |
| ... | ... |
| ... | ... |
| ... | ... |

- **Further classes are needed** → in some cases most likely an **iterative approach** necessary
- **Clauses may fit into multiple classes** → introduce **additional class** accordingly

Approach and First Steps | Building the Corpus



The labelling process and the information being collected:

- **Step 1: Manually collect the clauses** and the associated information from the Terms of Service agreements provided
- **Step 2: Label the collected data** according to the devised classification

| File Number | Company | Paragraph Title | Paragraph Text | Clause Title | Clause Text | Clause Label 1 | Clause Label 2 |
|-------------|---------|-----------------|----------------|--------------|-------------|----------------|----------------|
|-------------|---------|-----------------|----------------|--------------|-------------|----------------|----------------|

- Possibility to **collect further information about the companies** the Terms of Service agreements are used by, e.g., nature of the products being offered by the particular website

Outline of the Thesis | Overview

- Terms of Service agreements and classes being used
 - Brief **overview of Terms of Service agreements** and their **economic significance**
 - Description of the **classes that were used to divide the corpus**
- Approaches to **Text Classification**
- Overview of the **data used**
 - Information about the **companies** the clauses were obtained from, **distribution of classes** within the corpus, **etc.**
- Description of the **experimental approach**
 - **Models** used
 - **Observations**
 - **Results**
- **Analysis and conclusion**
 - **Discussion** of results and observations and their **possible causes**
 - **Potential future extensions** of the project



B.Sc. Computer Science

Jan Robin Geibel

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

Boltzmannstraße 3
85748 Garching bei München

Tel +49.89.289.
Fax +49.89.289.17136

www.matthes.in.tum.de



Appendix | References



[1] „What Is a Terms of Service Agreement?“, [upcounsel.com](https://www.upcounsel.com/terms-of-service-agreement), accessed on 12.06.2020

[2] „7 von 10 Internetnutzern in der EU kaufen online“, [Statistisches Bundesamt](https://www.destatis.de/DE/Home/Home.html), accessed on 12.06.2020

[3] Obar, J. A. and Oueldorf-Hirsch, A., 2020. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, 23 (1)

[4] Elshout, M., Elsen, M., Leenheer, J., Loos, M. and Luzak, J., 2016. Study on Consumers' Attitudes Towards terms Conditions (T&Cs) Final Report. *Report for the European Commission, Consumers, Health, Agriculture and Food Executive Agency (Chafea) on behalf of Directorate-General for Justice and Consumers.*

[5] „AGB-Check – AI-Supported Legal Review of terms and Conditions to Strengthen Consumer Protection“, [matthes.in.tum.de](https://www.matthes.in.tum.de/), accessed on 14.06.2020

[6] „Software Aided Analysis of Terms of Services“, [matthes.in.tum.de](https://www.matthes.in.tum.de/), accessed on 14.06.2020