

Eberhard Karls Universität Tübingen  
Mathematisch-Naturwissenschaftliche Fakultät  
Wilhelm-Schickard-Institut für Informatik

# LinguisticTextAnnotation

## **Specification**

Oliver Brehm

31.01.2017

# Inhaltsverzeichnis

|     |                                |   |
|-----|--------------------------------|---|
| 0.1 | Requirements . . . . .         | 3 |
| 0.2 | Software Komponenten . . . . . | 5 |
|     | 0.2.1 Überblick . . . . .      | 5 |
|     | 0.2.2 Komponenten . . . . .    | 5 |
| 0.3 | Datenmodell . . . . .          | 7 |
| 0.4 | User Interface . . . . .       | 9 |

## 0.1 Requirements

**Text Analyse** Der User möchte einen Flietext eingeben und von der Anwendung annotieren lassen. Nach der Eingabe soll das Ergebnis annotiert in der App dargestellt werden.

**Anpassung der annotierten Darstellung** Der User möchte die Parameter des annotierten Texts ändern. Angepasst werden können sollen Texteigenschaften (Font, Zeilenabstand), Darstellung der Annotation (Farben, Fonts für betonte und unbetonte Silbe), Zusätzliche Visualisierung für die Wortsegmentierung (Silbenbögen, Klammern, Abstände)

**Export der annotierten Darstellung** Der User hat die Möglichkeit verschiedene Formate des annotierten Texts zu exportieren, z.B. Druck, HTML oder Word.

**Verwaltung von User Accounts** Dem User soll die Möglichkeit gegeben werden, einen User Account zu erstellen, um persönlch verwendete Daten (z.B. Texte, Wortsegmentierungen) speichern zu können. Dazu müssen Funktionen und Interfaces für Registrieren eines Nutzeraccounts, Login, Logout, Bearbeiten der Nutzderinformationen und Löschen des Accounts bereitgestellt werden.

**Behandlung unbekannter Wörter** Dem User soll durch Klicken auf ein Wort oder einen Button "Unbekannte Wörter hinzufügen" die Möglichkeit gegeben werden, nacheinander die Segmentierung von Wörtern, die durch das System nicht eindeutig bestimmt wurden konnten, selbst festlegen zu können.

**Bestimmung der Segmentierung unbekannter Wörter** Für ein unbekanntes Wort soll in einem neuen View die segmentierung ausgewählt werden können. Dafür werden folgende Möglichkeiten gegeben:

- Segmentierungssystem aus der Bachelorarbeit von Leona Göbbels (CITE)
- Input aus G2P Systemen wie MARY
- Manuelle Segmentierung mit geeignetem User Interface

**Speicherung von Nutzer Segmentierungen** Vom Nutzer hinzugefügte Segmentierungen sollen (lokal für diesen Nutzer) gespeichert werden können und beim nächsten Vorkommen in einem Text automatisch verwendet werden.

**Speicherung von Annotationskonfigurationen** Die Einstellungen, die ein Nutzer an einenm annotierten Text vorgenommen hat, können als Vorlage für andere Texte gespeichert werden.

**Auswahl einer Annotationskonfiguration für einen Text** In den Annotationseinstellungen eines Textes kann eine zuvor gespeicherte Konfiguration verwendet werden.

**Speicherung von Nutzertexten** Analysierte Texte können vom Nutzer zusammen mit der verwendeten Konfiguration gespeichert werden. Den Texten können Metadaten zugeordnet werden, z.B. Thema, Niveau, Zielgruppe.

**Auflistung von Nutzertexten** Im Benutzerbereich werden die Texte, die der Nutzer hinzugefügt hat, geeignet strukturiert, dargestellt.

**expert user functions** TODO

## 0.2 Software Komponenten

### 0.2.1 Überblick

Grobstruktur der Komponenten:

- RESTful python backend
  - REST API
  - Dictionary Service -> Word DB
  - User Service -> User DB
- Web frontend: (AngularDart application)
  - TextAnnotation Component
  - User Component

### 0.2.2 Komponenten

Aufbau der Komponenten im Detail:

TODO



## 0.3 Datenmodell

Aus den Anforderungen lässt sich folgendes Datenmodell ableiten:

TODO





## 0.4 User Interface

TODO user interface designs for each requirement

