

Projektrealisierung

Von Lukas Bonn, Amina Uicker-Darwish,
Jan Rüdert, Aymane Bouguern

Agenda

- 1 Projektauftrag
- 2 Lasten- und Pflichtenheft
 - Funktionale- und nicht funktionale Anforderungen
 - Projektmanagement
 - Toolentwicklung
 - Datengrundlage
- 3 Projektstrukturplan
- 4 Herausforderungen
 - Bisherige Herausforderungen
 - Erwartete Herausforderungen
- 5 Ausblick

Projektauftrag



Auftraggeber, Projektausrichtung und Ausgangssituation



Projektziel und -inhalt



Projektabgrenzung und mögliche Risiken



Nutzen und Rahmenbedingungen

Auftraggeber, Projektausrichtung und Ausgangssituation

Projektname

- Entwicklung eines NLP-Tools zur Textzusammenfassung- und klassifikation

Auftraggeber

- Findet im Rahmen eines Hochschulprojekts statt
- Projektauftrag geht von den Dozierenden aus

Projektausrichtung

- Textverarbeitung
- Projektmanagement
- Forschungs- und Bildungsprojekt

Ausgangssituation

- Notwendigkeit eines leistungsfähigen NLP-Tools zur Textklassifikation und Zusammenfassung
- Manuelles Klassifizieren und Zusammenfassen von Texten kann zeitaufwendig, fehleranfällig und ineffizient sein

Projektziel/ Inhalt



Projektziel:

- Entwicklung eines NLP-Tools zur Textklassifikation und Zusammenfassung
- Frei wählbare Kompressionsrate bei der Zusammenfassung
- Benutzeranwendung



Projekthinhalt:

Planung, Organisation, Konzeption eines umfangreichen NLP-Entwicklungsprojekts

Entwicklung des eigentlichen NLP-Tools:

- Textklassifikation: Einteilung von Texten in verschiedene vordefinierte Kategorien
- Textzusammenfassung: Zusammenfassen von Texten unter Berücksichtigung einer Kompressionsrate
- Webanwendung als Schnittstelle zu dem Benutzer inkl. Eingabe einer Kompressionsrate

Projektabgrenzung und mögliche Risiken

Projektabgrenzung

- Sprachliche Begrenzung
- Textarten
- Kategorien

Mögliche Risiken und Gegenmaßnahmen

- Unzureichende Textdaten für die Klassifikation und Zusammenfassung
- Komplexität der NLP-Algorithmen und Herausforderungen bei der Modelloptimierung
- Keine rechtzeitige Lieferung des Tools mit den dazugehörigen Komponenten
- Unzureichende Dokumentation des entwickelten Tools und der Algorithmen

Risiken bei Nichtdurchführung des Projekts

- Verbleibende Ineffizienz manueller Prozesse
- Begrenzte Skalierbarkeit

Nutzen und Rahmenbedingungen

Nutzen des Projekts

- Zeitersparnis
- Steigerung der Produktivität
- Kosteneinsparungen

Voraussetzung / Rahmenbedingungen / Abhängigkeiten

- Verfügbarkeit von Textdaten
- Zugriff auf NLP-Methoden und –Algorithmen
- Expertise in maschinellem Lernen:

Wesentliche beteiligte Abteilungen / Bereiche

- Forschungsprojekt der Hochschule
- Studenten hauptsächlich für die Umsetzung verantwortlich
- Daher sind keine weiteren Personen in dem klassischen Sinne mit in die Entwicklung involviert

Lasten- und Pflichtenheft



- Funktionale Anforderungen
- Nicht-funktionale Anforderungen

Lastenheft

- Projektmanagement
- Textzusammenfassung
- Textklassifikation
- Webanwendung
- Datengrundlage

Pflichtenheft

Lastenheft: Funktionale Anforderungen



Zusammenfassungsfunktion

Umsetzung der gewünschte Kompressionsrate (z.B. 80%)



Klassifikationsfunktion

Klassifikation von Nachrichtenartikeln in englischer Sprache in verschiedene Oberkategorien



Bereitstellung einer Benutzerschnittstelle

Entgegennahme der Texte und Möglichkeit für den Benutzer, gewünschte Kompressionsrate auszuwählen



Mehrstufiger Ansatz

Verfolgung eines mehrstufigen Ansatzes zur Realisierung



Eigenständige Auswahl der Toolchain

Eigene Auswahl der Tools basierend auf den spezifischen Anforderungen

Keine fertigen Tools oder Lösungen verwenden

Lastenheft: Nicht-funktionale Anforderungen



Leistung:

Ausführung der Funktionen in angemessener Zeit



Benutzerfreundlichkeit:

Einfache und unkomplizierte Handhabung
Möglichst weitgehender Verzicht auf Schulungen
oder Anleitungen



Nachvollziehbarkeit:

Einfache Gestaltung des Entwicklungsprozesses
Auftraggeber kann den Prozess besser
nachvollziehen



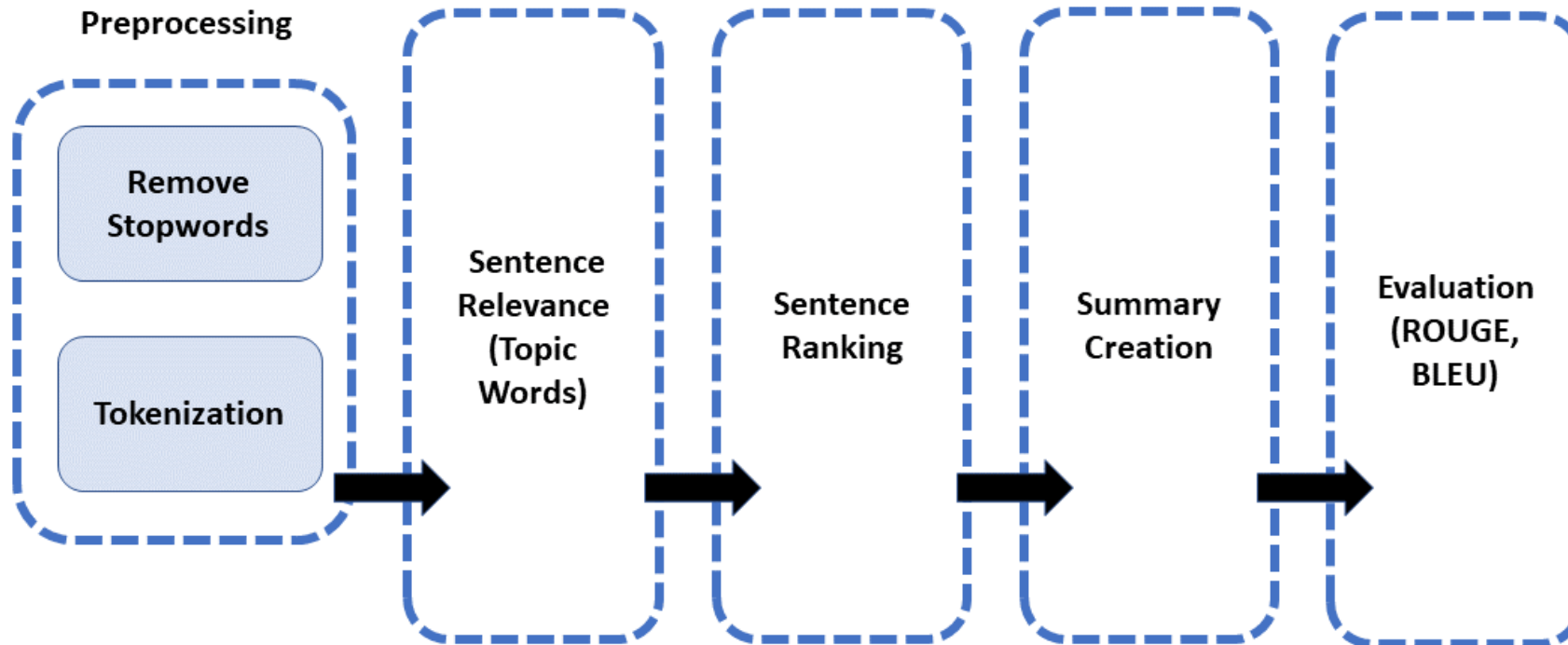
Projektmanagement - Projektorganisation

- Vier Projektmitglieder
- Projektleiter für die Kommunikation mit den Stakeholdern
- Zu gleichen Teilen an der Umsetzung beteiligt
- Genauere Aufgabeneinteilung erfolgt im späteren Projektverlauf
- Klassische Projektmanagementmethode

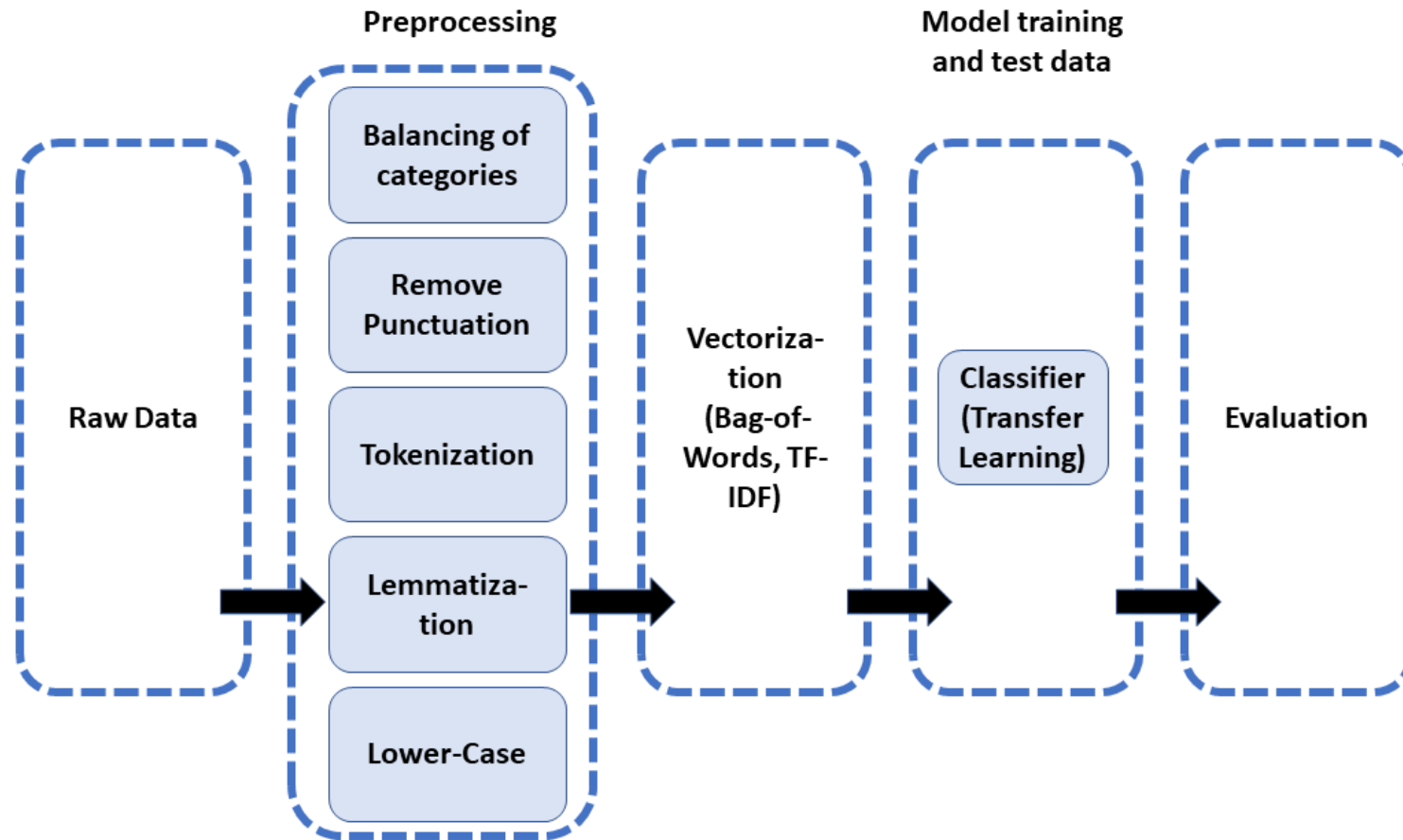
Projektmanagement - Meilensteine

Abgabefristen / Meilensteine	Laufzeit / Endtermin	Beschreibung
Meilenstein 1: Erste Planung	24.05.2023	Formulieren des Projektauftrags, Erstellen des Lasten- und Pflichtenheft und Entwerfen des groben Projektstrukturplans
Meilenstein 2: Detaillierte Planung	22.06.2023	Entwerfen des detaillierten Projektstrukturplans und Gantt-Charts
Meilenstein 3: Toolentwicklung und Abschlussprojektbe- richterstellung	27.07.2023	Entwicklung der Textzusammenfassung, -klassifikation und Website. Erstellen des Projektabschlussberichts, Erarbeitung der Lessons Learned und Durchführung einer kritischen Reflexion

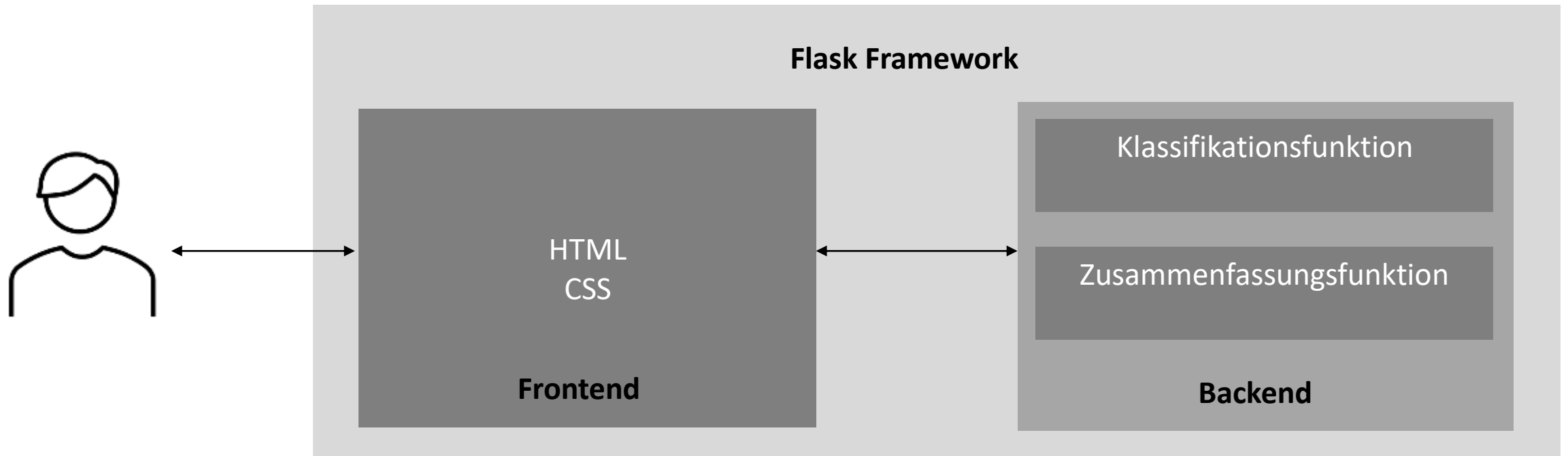
Textzusammenfassung - Pipeline



Textklassifikation - Pipeline



Webanwendung

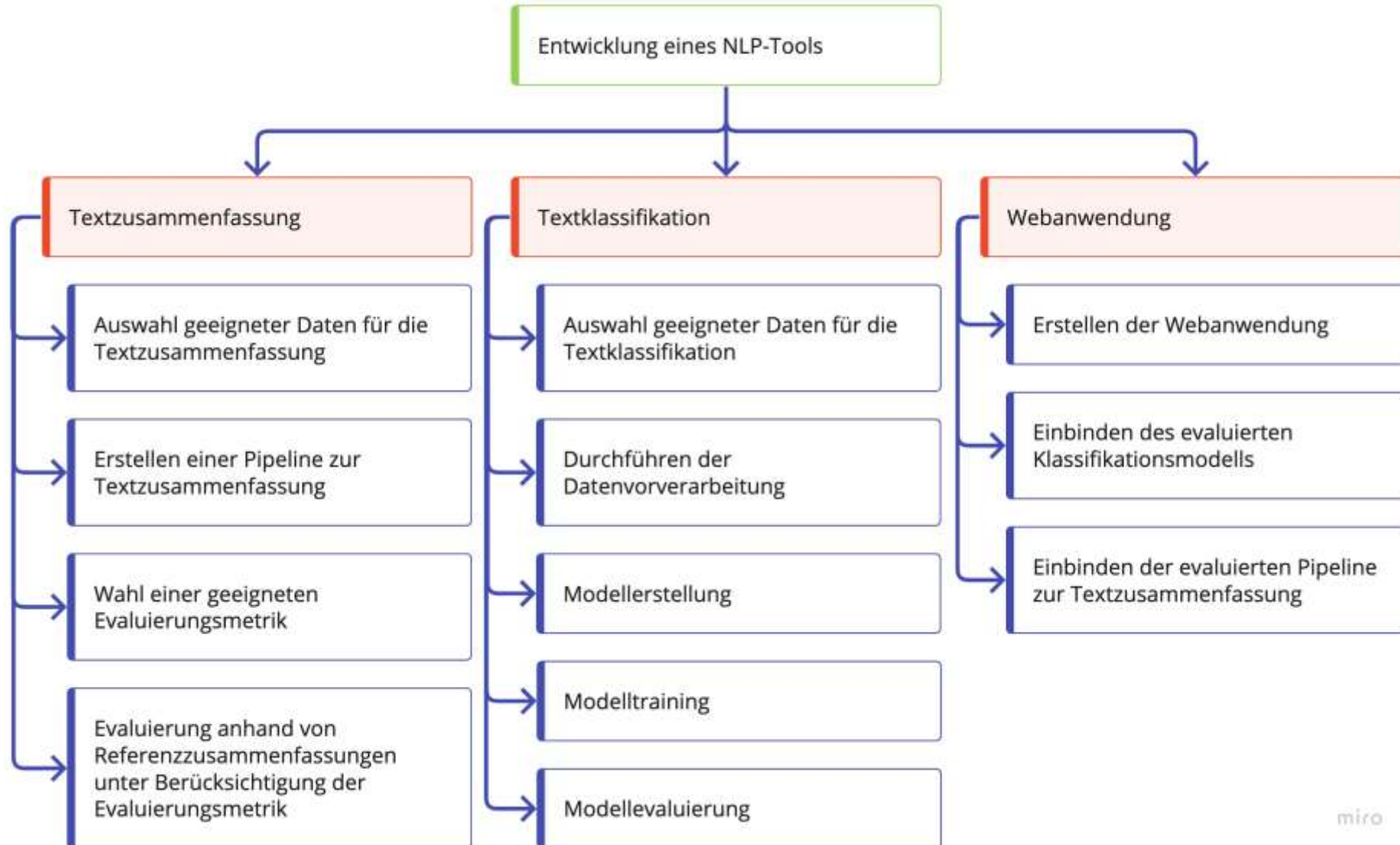


Datengrundlage

- Textzusammenfassung:
 - Ausreichende Länge der Texte
 - Referenzzusammenfassungen müssen "extractive" erstellt worden sein
- Textklassifikation:
 - Annotierter Datensatz
 - Balanced
 - Klassifikation von Artikeln vorgesehen



Projektstrukturplan





Bisherige Herausforderungen

- Datenbeschaffung:
Finden eines geeigneten Datensatzes sowohl für Klassifikation als auch für Zusammenfassung
- Auswahl eines Ansatzes für die Textzusammenfassung unter Berücksichtigung der Kompressionsrate

Erwartete Herausforderungen

- Transfer Learning aufgrund fehlender Erfahrung in diesem Bereich
 - Zeitliche Herausforderung die Deadlines zu treffen
 - Einbinden des evaluierten Modells in die Webanwendung
 - Modellintegration
 - Einbinden der Textzusammenfassungsfunktion
 - Abhängigkeiten der einzelnen Arbeitspakete
- > Erschwerte Zusammenarbeit



Ausblick

- Wahl eines konkreten Datensatzes
- Nähere Definition und Beschreibung der Arbeitspakete (Verwendete Hilfsmittel, Voraussetzungen, Beteiligte Personen, etc.)
- Zuordnung zeitlicher Aufwände für die definierten Arbeitspakete
- Visualisierung in einem Gantt-Chart