**Projektbericht zum Modul Information Retrival und**

**Visualisierung Sommersemester 2021**

**Titel des Dokuments**

Richard Brennecke

Matrikelnummer:

# Inhaltsverzeichnis

[1. Inhaltsverzeichnis 2](#_Toc79327777)

[2. Einleitung 3](#_Toc79327778)

[2.1 Anwendungshintergrund 3](#_Toc79327779)

[2.2 Zielgruppen 3](#_Toc79327780)

[2.3 Überblick und Beiträge 3](#_Toc79327781)

[3. Daten 4](#_Toc79327782)

[3.1 Technische Breitstellung der Daten 4](#_Toc79327783)

[3.2 Datenvorverarbeitung 4](#_Toc79327784)

[4. Visualisierung 5](#_Toc79327785)

[4.1 Analyse der Anwendungsfälle 5](#_Toc79327786)

[4.2 Anforderungen an die Visualisierungen 5](#_Toc79327787)

[4.3 Präsentation der Visualisierung 5](#_Toc79327788)

[4.3.1 Visualisierung Eins 5](#_Toc79327789)

[4.3.2 Visualisierung Zwei 5](#_Toc79327790)

[4.3.3 Visualisierung Drei 5](#_Toc79327791)

[4.4 Interaktion 5](#_Toc79327792)

[5. Implementierung 5](#_Toc79327793)

[6. Anwendungsfälle 6](#_Toc79327794)

[6.1 Anwendung Visualisierung Eins 6](#_Toc79327795)

[6.2 Anwendung Visualisierung Zwei 6](#_Toc79327796)

[6.3 Anwendung Visualisierung Drei 6](#_Toc79327797)

[7. Verwandte Arbeiten 6](#_Toc79327798)

[8. Zusammenfassung und Ausblick 6](#_Toc79327799)

[9. Anhang 7](#_Toc79327800)

# Einleitung

* Problem:
  + Analyse von verschiedenen Wein Daten -> Interessante Zusammenhänge herausfinden
* Fragen welche Beantwortet werden sollen
  + Gibt es zusammenhänge zwischen den verschiedenen Daten?
    - Scatterplot
  + Hängen Daten über mehrere Dimensionen zusammen?
    - Parallele Koordinaten
  + Wo kommen die meisten Weine her?
    - Baumhierarchie

## Anwendungshintergrund

* Erklärung der Informationsvisualisierungen
* Kleine Beantwortung der Fragen mithilfe der Visualisierung

## Zielgruppen

* Weininteressierte
  + Vorwissen
    - Kaum bis gar nicht
  + Erkenntnisse:
    - Zusammenhang von verschiedenen Kriterien beim Wein
      * Zusammenhänge kurz erklären
    - Informationsgewinnung
      * Kennenlernen von Weinen
    - Entdeckung neuer Weine welche sie trinken möchten
* Weineinkäufer
  + Vorwissen
    - Vorhanden bis Exzellent
  + Erkenntnisse:
    - Entdecken von neuen Sorten die sein Sortiment ergänzen
    - Entdeckung von neuen Sorten die ggf. Außergewöhnlich sind
    - Beratung der Kunden die gewisse Vorlieben haben
* Weinexperte
  + Vorwissen
    - Gut bis ausgeprägt
  + Erkenntnisse:
    - Entdecken von neuen Sorten die seinem Geschmack entsprechen
    - Bessere Einschätzung seiner bisherigen Weine

## Überblick und Beiträge

* Erklären welche Daten verwendet wurden
  + Kurz auf Aufbereitung eingehen
* Visualisierungstechniken erklären
* Beiträge???

# Daten

* Beschreibung der gegebenen Daten
* Eignung der Daten für die Zielgruppen
  + Weintressierte
    - Gut
  + Weinexperte/ Weinverkäufer
    - Teilweise
    - Daten können unvollständig sein -> und haben zu wenig Aussagekraft mit Body, Süße usw.
* Fragestellungen
  + Gut da Dimensionen erkannt werden können
  + Herkunft der Daten erkennbar -> Teilweise aber unvollständig
* Daten Ergänzung
  + Mussten mit Geo Daten ergänzt werden für Baumhierarchie da sonst kein Ursprungsknoten

## Technische Breitstellung der Daten

* Daten Zugänglich?
  + Sind über GitHub für die verschiedenen Darstellungen erreichbar
* Formate
  + CSV -> Für alle Daten die nicht ergänzt wurden
  + JSON -> Für alle Geo Daten
* Besonderheiten
  + CSV
    - 0 oder nichts bedeutet dort ist nichts vorhanden
    - Trennung durch normales komma
  + JSON
    - Nur Name und Beziehung (Eltern – Kind) in der Datei vorhanden
    - Länder welche keine Weine Produzieren wurden außen vor gelassen

## Datenvorverarbeitung

* Datenverarbeitungsschritte
  + Sichten der Daten (Umwandeln in besser Lesbares Format -> Excel)
  + Bearbeiten der Daten
    - Durchschnitte Bilden
    - Namen überarbeiten
    - Übersetzten
  + Rücküberführung zur CSV
* Daten weggelassen
  + Daten nicht mehr lesbar (eine Japanischer Wein wo nicht mehr Rückschlüsse gezogen werden könnten)
* Durschnitte
  + Gebildet über die Trinktemperatur, Alkoholgehalt
  + Daten konnten ansonsten nicht eingelesen werden
  + Außerdem lag der unterschied durchschnittlich nicht bei mehr als 2 Gard
* Aussagekräftiger?
  + So hat man noch Toleranz beim der Temperatur und Alkohol ohne das sich die anderen Werte ändern müssten (Body, Süße usw.)

# Visualisierung

* Analyse kann erst gemacht werden, wenn Visualisierungen fertig sind

## Analyse der Anwendungsfälle

* Analyse kann erst gemacht werden, wenn Visualisierungen fertig sind

## Anforderungen an die Visualisierungen

* Analyse kann erst gemacht werden, wenn Visualisierungen fertig sind

## Präsentation der Visualisierung

* Analyse kann erst gemacht werden, wenn Visualisierungen fertig sind

### Visualisierung Eins

* Wird ein Scatterplot

### Visualisierung Zwei

* Wird Parallele Koordinaten

### Visualisierung Drei

* Wird eine Baumhierarchie

## Interaktion

* Scatterplot und Parallele Koordinaten
  + Buttons zum verändern/ verschieben der Dimensionen
* Baumhierarchie
  + Keine nur anschauen

# Implementierung

* Kann erst eingeschätzt werden, nachdem es fertig gestellt worden ist
  + Aktuell hoher Aufwand und nur Baumhierarchie konnte sehr einfach aus Übung übernommen werden

# Anwendungsfälle

* Erst nach Fertigstellung der Visualisierungen möglich

## Anwendung Visualisierung Eins

* Anwendungsfall für Scatterplot

## Anwendung Visualisierung Zwei

* Anwendungsfall für Parallele Koordinaten

## Anwendung Visualisierung Drei

* Anwendungsfall für Baumhierarchie

# Verwandte Arbeiten

* Aktuell noch nicht recherchiert

# Zusammenfassung und Ausblick

* Ausblick er bei fertigem Projekt möglich

# Anhang