**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 4](#_Toc79327781)

[3. Daten 4](#_Toc79327782)

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

[3.2 Datenvorverarbeitung 4](#_Toc79327784)

[4. Visualisierung 4](#_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 6](#_Toc79327789)

[4.3.2 Visualisierung Zwei 6](#_Toc79327790)

[4.3.3 Visualisierung Drei 6](#_Toc79327791)

[4.4 Interaktion 6](#_Toc79327792)

[5. Implementierung 6](#_Toc79327793)

[6. Anwendungsfälle 7](#_Toc79327794)

[6.1 Anwendung Visualisierung Eins 7](#_Toc79327795)

[6.2 Anwendung Visualisierung Zwei 7](#_Toc79327796)

[6.3 Anwendung Visualisierung Drei 7](#_Toc79327797)

[7. Verwandte Arbeiten 7](#_Toc79327798)

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

[9. Anhang 10](#_Toc79327800)

Seitenabstand Ränder alle 2

Schriftart: Times new roman

Zeilenabstand: 1,5

Zitierung: mit eckige Klammer mit Nummer dahinter

# Einleitung

* Zielproblem:
  + Analyse von verschiedenen Wein Daten
    - Interessante Zusammenhänge herausfinden
    - Erkenntnisgewinn aus diesen Daten
* Fragen welche Beantwortet werden sollen
  + Gibt es Zusammenhänge zwischen dem Körper/Body/Süße/Alk. und dem Preis eines Weines?
    - Scatterplot
  + Hängen Daten über mehrere Dimensionen zusammen?
    - Parallele Koordinaten
  + Wo kommen die meisten Weine her?
    - Baumhierarchie

## Anwendungshintergrund

* Erklärung der Informationsvisualisierungen
  + Was ist ein Scatterplott/ parallele Koordinaten/ Baumhierachie?
* Hintergrund Daten bereitstellen?
  + Informationen zu verschiedenen Daten
    - Wie kann der Körper/ Süße/ Säure/ Gerbstoffe bestimmt werden?
    - Was haben die Jahre für Einfluss auf die Weine?
    - Was bedeuten die Felder Verwendung und Type (Sorte, Lokal, Produzent)

## 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
  + Oberkategorien kurz erklären
* Visualisierungstechniken erklären
* Beiträge
  + Mehrwert der Techniken für die Darstellung der Daten
    - Sacatterplott -> Gegenüberstellung von Dimensionen

# 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
  + Und wo kommen die her?

## 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
  + Bekanntmachen mit den Daten (Umwandeln in besser Lesbares Format -> Excel)
  + Bearbeitung der Daten
    - Namen überarbeiten (aus den Zahlen entfernen)
    - Zahlen bearbeiten
      * Umrechnung von WON in Euro
      * Durchschnitte Bilden
    - Namen werden Überarbeitet (Umlaute und Apostrophe nicht richtig konvertiert)
    - Übersetzten der Spaltenüberschriften
    - Bereitstellung der Daten für JSON
    - Herauslösen der Datensätze, welche nicht komplett sind
  + Überführung der Daten
* Daten weggelassen
  + Daten nicht mehr lesbar (eine Japanischer Wein wo nicht mehr Rückschlüsse gezogen werden könnten)
  + Nichts Werte
* 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 dass sich die anderen Werte ändern müssten (Body, Süße usw.)

# Visualisierung

* Analyse kann erst gemacht werden, wenn Visualisierungen fertig sind

## Analyse der Anwendungsaufgaben

* Analyse kann erst gemacht werden, wenn Visualisierungen fertig sind
* Anwendungsaufgaben
  + Wie helfen die Darstellungen die genannten Problemstellungen zu beantworten?
  + Hauptziel möglich viele Einblicke in das Thema der Weine zu erhalten
  + Zusammenhänge zwischen den einzelnen Eigenschaften
  + Zusammenhang von Produktionsmenge eines Landes und der Anzahl der Weine
  + Hervorstechen von Datensätze aus Allgemeinheit, Zusammenhänge oder Trends von Eigenschaften
* Mentale Modelle
  + Welche Visualisieren eigenen sich um das alles zu kombinieren von Wissen und den Modellen (intuitiv erkennbar)

## 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
* Vorstellen, Interaktivität, Designentschiedungen begründen Diskutieren wieso nicht anderen Techniken verwenden worden sind

### Visualisierung Eins

* Wird ein Scatterplot
* Präsentation -> Abbildung, Kodierung der Daten, Interaktionsmöglichkeiten
* Erfüllung und wie gut die Anforderungen erfüllt werden
* Warum ist die Visuelle Darstellung passend für das Problem? (Diskussion der Auswahl von Darstellungen)

### Visualisierung Zwei

* Wird Parallele Koordinaten
* Präsentation -> Abbildung, Kodierung der Daten, Interaktionsmöglichkeiten
* Erfüllung und wie gut die Anforderungen erfüllt werden
* Warum ist die Visuelle Darstellung passend für das Problem? (Diskussion der Auswahl von Darstellungen)

### Visualisierung Drei

* Wird eine Baumhierarchie
* Präsentation -> Abbildung, Kodierung der Daten, Interaktionsmöglichkeiten
* Erfüllung und wie gut die Anforderungen erfüllt werden
* Warum ist die Visuelle Darstellung passend für das Problem? (Diskussion der Auswahl von Darstellungen)

## Interaktion

* Scatterplot und Parallele Koordinaten
  + Buttons zum verändern/ verschieben der Dimensionen
* Baumhierarchie
  + Keine nur anschauen
* Zweck der Interaktion
* Warum wurden andere Interaktionen umgesetzt und nicht andere?
* Begründung Interaktion zwischen denen nicht mit dabei

# Implementierung

* Kann erst eingeschätzt werden, nachdem es fertig gestellt worden ist
  + Aktuell hoher Aufwand und nur Baumhierarchie konnte sehr einfach aus Übung übernommen werden
* Gliederung des ELM Codes
* Übungsadaption
* Datenstruktur Modells bei den verschiedenen Interaktionen
* Bei uns in einem Record gespeichert im Main und dann im Update wird auf einen record zugegriffen

# Anwendungsfälle

* Erst nach Fertigstellung der Visualisierungen möglich
* Spezifischer Anwedungsfall -> wo Muster da sind oder nicht was es zu was besonderen macht
* Relevanz für die Zielgruppe
* Möglichkeit Umsetzung mit anderen Personen

## Anwendung Visualisierung Eins

* Anwendungsfall für Scatterplot
* Preis und Körper -> Je höher der Preis desto höher kann der Körper des Weines sein. -> Es muss aber nicht immer ganz teuer sein
* Teuersten Weine haben tendenziell einen größeren Körper

## Anwendung Visualisierung Zwei

* Anwendungsfall für Parallele Koordinaten
* Zusammenhang von Körper, Gerbstoffe, Süße, Säuregehalt
* Keine Erkennbares Muster -> Bedeutet weine können je nach Wunsch des Kunden unterschiedliche Schmecken
* Nur Süße 5 kann maximal 4 Säure haben und Säure 5 nur maximal 3 Süße
* Viele Hohe gerbstoffe 5 maximal 2 in süße
* Ansonsten gehen alle Gerbstoffe auch auf die 5 in Süße
* Niederoger Körper hat wenig Gerbstoffe -> Hinweis auf Zusammenhang dieser

## Anwendung Visualisierung Drei

* Anwendungsfall für Baumhierarchie
* Heraussuchen verschiedener Asiatischer Weine
* Georgien hat die meisten Weine
* Israel die wenigsten
* Japan und Süd-Korea gleich viele

# Verwandte Arbeiten

* Aktuell noch nicht recherchiert
* Zwei Artikel diskutieren
  + Gemeinsamkeiten und Unterschiede dabei herausstellen
  + Visualisierung von Krankenhausdaten
  + <https://www.nm.informatik.uni-muenchen.de/common/pub/Fopras/petr02/PDF-Version/petr02.pdf>
  + Daten zur Multiplen Sklerose (S.7)
  + Unterschiede
    - Vorherrein Berechnung von statistischen Größen (Median, Normalverteilung) (s.10)
    - Dateneinteilung mithilfe von Klassen -> dient der Übersichtlichkeit (s.10)
    - Datenaufbereitung erfolgt mithilfe von ColdFusion (s.10) und die Datendarstellung (s.24)
    - Drop Down Liste der Benutzeroberfläche (für x und y Achse) (s.38)
    - Visualisierung und Interaktion bezieht sich auf ein Balkendiagramm, welches durch weitere Interaktionen verändert werden kann (s.39)
    - Scatterplot nur das Ergebnis, welcher aber um Konfidenzintervalle und Regressionsgerade erweitert (s.41)
  + Gemeinsamkeiten
    - Einfache Benutzeroberfläche welche auch Auswahl der Attribute zulässt (s.11)
    - Verwendung des Scatterplots als Darstellung (s.11)
  + Visualisierung von Patientendaten mithilfe eines Programms
  + <https://repositum.tuwien.at/bitstream/20.500.12708/4082/2/Fels%20Ulrich%20-%202015%20-%20Usability%20Analyse%20des%20Programms%20Animated%20Scatter%20Plot.pdf>
  + Daten zur Diabetespatienten (s. V) (Daten mit Zeitorientierung)
  + Unterschiede
    - Analyse über die Nutzbarkeit des Scatterplots (s. V)
    - Animierung des Scatterplots über die zeitreihen hinweg bedient erfolgt dabei ähnlich die eines Videorecorders (s.55)
    - Auswahl der Achsen ist möglich (s.53) über drop down Menüs (s.54)
    - Anpassungen der Darstellung in Größe, Form, Farbe (s.58)
      * Anpassung von Größe und Form anhand von Attributen (s.59)
    - Verschieden Grenzwerte können angezeigt werden und sind individuell anpassbar (s.57)
    - Nachverfolg der Daten mithilfe von Spuren (zeitliche) (s.58)
    - Möglichkeit des herein Zoomens in die Abbildung (s.60)
    - Daten können gefiltert werden nach wünschen des Benutzers (s.60)
    - „Berühren“ der Maus eines Punktes gibt genaue Werte der Untersuchung wieder (inkl. Spurennachverfolgung) (s.62)
  + Gemeinsamkeiten
    - Darstellung der Daten über einen Scatterplot (s. V)

# Zusammenfassung und Ausblick

* Ausblick er bei fertigem Projekt möglich
* Zusammenfassung der Beiträge
* Mehrwert für Zielgruppe und Personen
* Erweiterungen für Ebene und Datenebene

Anhang