

Gesprächsbasiertes System für Weinempfehlung

Zusammenfassung

Thursday 18th June, 2020 - 15:12

Flavio De Jesus Matias
Universität Luxemburg
Email: flavio.dejesus.001@student.uni.lu

Sviatlana Höhn
Universität Luxemburg
Email: sviatlana.hoehn@uni.lu

Abstract—Dieses Dokument ist die Zusammenfassung des Bachelor-Semesterprojekts von Flavio De Jesus Matias unter der Leitung von Sviatlana Höhn.

1. Einführung & Projektbeschreibung

Chatbots sind heutzutage ein sehr wichtiges Werkzeug für die meisten Online-Unternehmen, da sie sich bei der Bearbeitung grundlegender Benutzeranfragen auf ihre virtuellen Assistenten verlassen. Chatbots präsentieren nur eine geringe Menge an Textdaten, um die Überlastung des Benutzers zu vermeiden. Die Notwendigkeit eines effektiven und automatischen Zusammenfassungssystems hat daher zugenommen. Daher ist es wichtig, über leistungsfähige Algorithmen zu verfügen, um viele Textdaten zu analysieren und zu kleinen lesbaren Zusammenfassungen zu verarbeiten.

Ziel des Projekts ist es, einen Algorithmus zum Zusammenfassen von Weinbewertungen zu entwickeln. Das Weinempfehlungssystem ruft dann die Zusammenfassungen auf Benutzerwunsch ab und zeigt sie an.

2. Voraussetzungen

2.1. Zusammenfassung aus Weinbewertungen generieren

Um richtig zu funktionieren, muss der Algorithmus in der Lage sein, beliebig viele Bewertungen zu verarbeiten und in eine Zusammenfassung umzuwandeln. Jede Zusammenfassung muss gemäß den Anforderungen des Chatbots erstellt werden, da sie später im Chatbot verwendet werden.

2.2. Natural language understanding & retrieval

Das NLU-Modell muss die verschiedenen Weinsorten erkennen und in der Lage sein, die richtigen Entitätswerte aus der Benutzeranforderung zu extrahieren. Dies gilt auch für die Anfragen der Weinsorten, die nicht in den Trainingsbeispielen sind. Aus den erkannten Entitätswerten muss der Chatbot die entsprechende Zusammenfassung abrufen und an den Benutzer senden.

3. Design & Produktion

3.1. Scientific Deliverable

Der wissenschaftliche Teil des Projekts besteht in der Entwicklung eines Algorithmus um Zusammenfassungen aus Weinbewertungen zu generieren. Der Algorithmus wurde getrennt und in 4 verschiedene Dateien organisiert (siehe Abbildung 1).

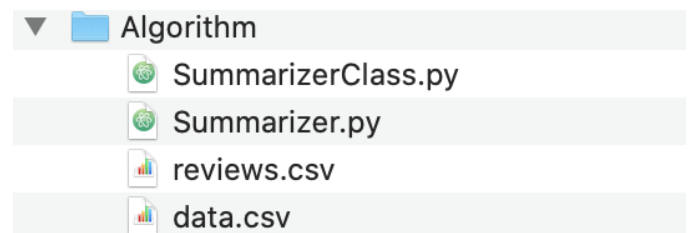


Figure 1: Dateistruktur des Algorithmus (4 Dateien)

3.1.1. data.csv. Diese Datei enthält die Anfangsdaten, bzw. Weinbewertungen, das aus ungefähr 130 Tausend Zeilen besteht.

3.1.2. SummarizerClass.py. Diese Datei erledigt alle Zusammenfassungen. Alle Bewertungen werden in Sätze unterteilt und diese dann bewertet. Es gibt **zwei** entscheidende Aspekte, wenn es um die Bewertung der Sätze geht: die Satzlänge und die Tf-idf-Punktzahl die doppelt zählt. Nachdem wir nun eine Liste mit allen Sätzen und der Punktzahl haben, können wir die endgültige Zusammenfassung erstellen, indem wir die n Sätzen mit der besten Punktzahl in dieser Liste nehmen.

3.1.3. Summarizer.py. Diese Datei ladet die Anfangs-Weinbewertungen und entfernt alle Duplikate sowie alle Weinsorten, die weniger als 20 Bewertungen haben. Der Datensatz, mit dem wir arbeiten, ist in Abbildung 2 dargestellt.

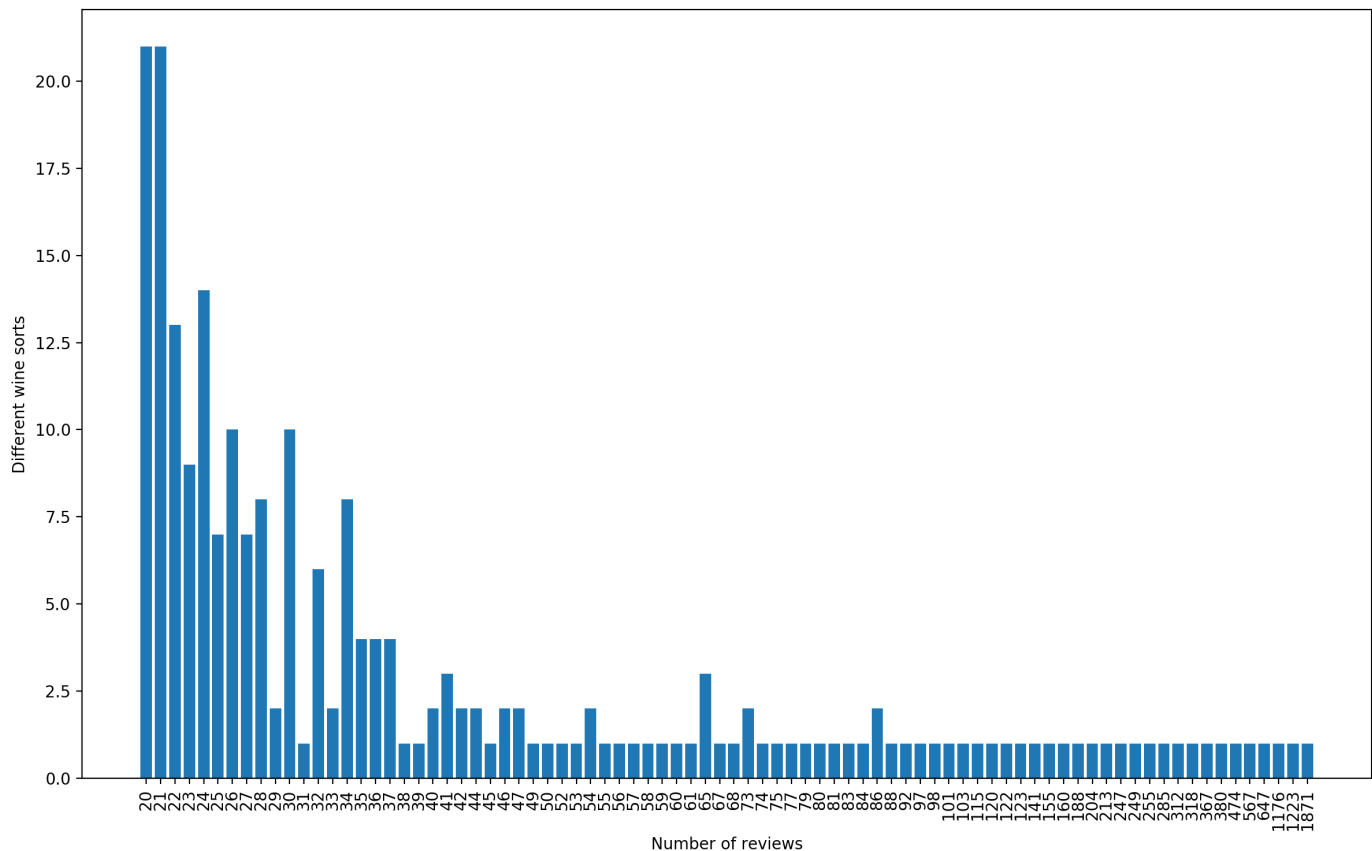


Figure 2: Anzahl der Bewertungen pro Weinsorten

3.1.4. reviews.csv. Diese Datei enthält alle generierte Zusammenfassungen und ist die Datenbank für den Chatbot.

3.2. Technical Deliverable

Das technische Teil des Projekts besteht darin, einen gesprächsbasiertes System für Weinempfehlung, bzw. einen Chatbot zu erstellen.

Der Chatbot wurde mit Rasa erstellt, einem Open-Source-Framework für KI-Chatbots. Nach dem Einrichten des Standard-Chatbots wurde jeder unnötige Teil des Codes entfernt. Als Erstes wurde die Domäne des Chatbots festgelegt, indem eine neue Absicht (Weinname) und neue Aktionen (um Informationen über einen Wein anzufragen) hinzugefügt wurden. Zweitens wurden Geschichten hinzugefügt, die echte Konversationsdaten zwischen einem Benutzer und dem Chatbot darstellen. Drittens musste das NLU-Modell trainiert werden. Für jede Weinsorte wurden 7 verschiedene Satz- und Fragevorlagen hinzugefügt. Dann wurde jedes Beispiel zweimal hinzugefügt: mit großgeschriebenen Anfangsbuchstaben und mit Kleinbuchstaben. Dies deckt die gesamte Datenbank ab. Da wir 227 verschiedene Weinsorten haben, entspricht dies $7 \cdot 2 \cdot 227 = 3178$ Trainingsbeispielen im NLU-Modell. Zum schluss

wurde eine neue Aktion hinzugefügt, um die Zusammenfassungen aus der Datenbank abzurufen und dem Benutzer anzuzeigen.

4. Bewertung & Abschluss

Die zu Beginn gestellten Anforderungen wurden erfolgreich erfüllt, da der Algorithmus die Zusammenfassungen erfolgreich generiert. Der Chatbot kann auch benutzerdefinierte Entitätswerte erkennen die nicht in den Trainingsbeispielen enthalten sind und den Benutzer mit der gewünschten Zusammenfassung beliefern.

In Zukunft könnte das Projekt durch das Hinzufügen von Informationen über die verschiedenen Weinsorten erweitert werden. Dies ermöglicht dem Benutzer mehr über seinem gewünschten Wein zu erfahren.