

Sprache, Spracheingabe, Text & Übersetzung mit Google Cloud APIs

M.Sc. Onur Yilmaz

Angewandte Künstliche Intelligenz

Schriftliche Ausarbeitung - Cloud Computing
Fachhochschule Südwestfalen

Gutachter: Prof. Dr. Giefers

10. August 2023

Inhaltsverzeichnis

1	Grundlagen	3
1.1	Einführung in Cloud Computing	3
1.1.1	Was ist eine API und ihre Funktion?	3
1.1.2	Bedeutung von API-Keys	3
1.2	Erstellen und Aufrufen der API-Anfrage	4
1.2.1	Grundlagen des API-Aufrufs	4
1.2.2	Spezifikationen und Parameter	5
1.3	Interaktion mit der Google Cloud API	6
1.3.1	Umgang mit dem Google Cloud Dienst	6
1.3.2	Erstellen eines API-Keys in Google Cloud	6
2	Spracherkennung und -transkription	7
2.1	Schritte im Speech-to-Text Prozess	7
2.2	Google Cloud Speech API	7
3	Sprachübersetzung	8
3.1	Erkennung und Übersetzung von Texten	8
3.2	Cloud Translation API	8
4	Textanalyse	9
4.1	Klassifizierung von Text in Kategorien	9
4.2	Entitäten- und Sentimentanalyse	9
4.3	Natural Language API	9
5	Sprachsynthese	10
5.1	Erzeugung synthetischer Sprache	10
5.2	Natural Language API	10

Einleitung

In dieser Arbeit werden Technologien und Anwendungen aus dem Bereich der Sprach- und Textverarbeitung beleuchtet, die auf Google Cloud-Diensten aufbauen. Hierzu gehören die *Cloud Speech API*, die *Cloud Translation API*, die *Natural Language API* und die *Text-to-Speech API* [1].

Im Abschnitt über die *Grundlagen* wird eine Einführung in die Konzepte des Cloud Computings, der API und des Umgangs mit dem Google Cloud Dienst gegeben.

Der Abschnitt *Spracherkennung und -transkription* fokussiert sich auf die *Cloud Speech API*, die die Transkription von Audio in Text ermöglicht, und die Methoden zur Messung und Verbesserung der Sprachgenauigkeit.

In der *Sprachübersetzung* wird die *Cloud Translation API* behandelt, die den Prozess der Übersetzung von Texten in verschiedene Sprachen ermöglicht.

Der Bereich *Textanalyse* befasst sich mit der *Natural Language API*, die Techniken zur Klassifizierung von Text in Kategorien und zur Analyse von Entitäten und Sentiments bietet.

Im Abschnitt *Sprachsynthese* wird die *Text-to-Speech API* vorgestellt, die die Erzeugung synthetischer Sprache ermöglicht.

Die Arbeit dient nicht nur als theoretischer Überblick, sondern bietet auch praktische Einblicke und Anleitungen zur Verwendung dieser Tools. Dabei werden unterschiedliche Schwierigkeitsgrade und Themenbereiche abgedeckt, um einen umfassenden Einblick in die Möglichkeiten der Sprach- und Textverarbeitung mit Google Cloud zu bieten.

1 Grundlagen

1.1 Einführung in Cloud Computing

Cloud Computing bezeichnet das Angebot und die Nutzung von IT Infrastrukturen und Anwendungen, die nicht lokal auf dem eigenen Computer, sondern in einem Netzwerk, meistens dem Internet, ausgeführt werden. Durch Cloud Computing können Unternehmen, aber auch Einzelpersonen Speicherplatz, Rechenleistung oder Softwareanwendungen als Service über das Internet nutzen.

Eine der Hauptanwendungen von Cloud Computing in der heutigen Zeit ist die Datenverarbeitung. Besonders hervorzuheben ist hier die Verwendung von Cloud-Plattformen für Natural Language Processing (NLP) Anwendungen, da sie die Fähigkeit bieten, große Mengen an Textdaten effizient zu analysieren und zu verarbeiten.

1.1.1 Was ist eine API und ihre Funktion?

Wenn man über Cloud spricht, ist eine der Hauptinteraktionen die Verwendung von APIs.

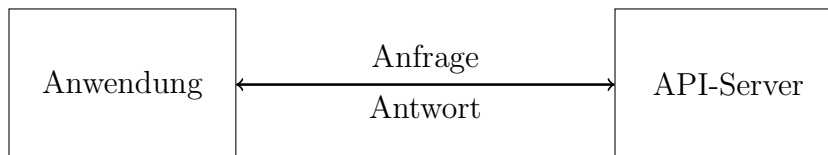
Eine API (Application Programming Interface) dient als Schnittstelle, die es Entwicklern ermöglicht, bestimmte Funktionen eines Programms oder einer Plattform zu nutzen, ohne sich mit den internen Details auseinandersetzen zu müssen.

1.1.2 Bedeutung von API-Keys

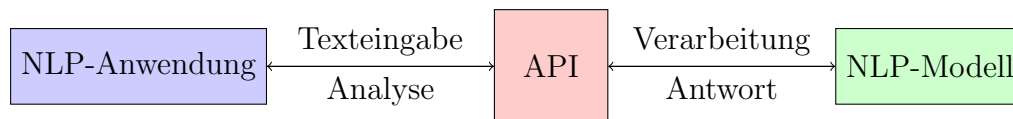
API-Keys sind ein wesentlicher Bestandteil vieler Webanwendungen und Cloud-Dienste. Sie dienen als Identifikator für den Benutzer oder Entwickler und stellen sicher, dass die Anfrage autorisiert ist. Durch den API-Key kann der Serviceanbieter den Zugriff auf seine Ressourcen kontrollieren und den Verbrauch überwachen.

1.2 Erstellen und Aufrufen der API-Anfrage

Das Aufrufen einer API, oft als „API-Anfrage“ bezeichnet (*im engl. request*), ist der Prozess, bei dem ein Programm oder eine Anwendung eine Anforderung an einen Server sendet und eine Antwort zurückerhält.



Im Kontext von NLP könnte eine API es ermöglichen, Textdaten zu senden und als Antwort eine Analyse oder Übersetzung dieses Textes zu erhalten.



1.2.1 Grundlagen des API-Aufrufs

Eine API-Anfrage ist eine Aufforderung an einen Server, eine bestimmte Aktion auszuführen oder Daten bereitzustellen.

Um zu verstehen, wie eine solche Anfrage aufgebaut ist, betrachten wir ihre Hauptbestandteile:

- i **Endpoint (URL)**: Dies ist die Webadresse, an die die Anfrage gesendet wird. Man kann es sich wie eine spezielle Telefonnummer vorstellen, die nur für bestimmte Dienste oder Informationen reserviert ist.
- ii **HTTP-Methode**: Dies bestimmt, welche Art von Aktion wir vom Server verlangen. Die gängigsten Methoden sind GET (um Daten abzurufen), POST (um neue Daten zu senden) und PUT (um bestehende Daten zu aktualisieren).
- iii **Header**: Zusätzliche Informationen, die mit der Anfrage gesendet werden. Hier können z.B. Anmeldedaten für den Server oder spezifische Anweisungen zur Datenverarbeitung enthalten sein.
- iv **Body**: Bei manchen Anfragen, wie z.B. POST oder PUT, schicken wir dem Server Daten im Body der Anfrage.

1.2.2 Spezifikationen und Parameter

Jeder Server oder Dienst kann mehrere Endpunkte haben, die unterschiedliche Arten von Daten oder Funktionen bedienen. Um die richtige Information oder Funktion zu erreichen, benötigen wir oft zusätzliche Details oder Parameter. Diese Parameter helfen dem Server, unsere Anfrage besser zu verstehen und die gewünschten Daten oder Dienste bereitzustellen.

Zum Beispiel, wenn wir Daten über einen bestimmten Benutzer von einer API abrufen möchten, könnten wir eine Benutzer-ID als Parameter in der URL oder im Body der Anfrage übergeben.

```
import requests

url = "https://api.example.com/userinfo"

headers = {
    "Authorization": "auth_token_here"
}

params = {
    "benutzerID": "12345"
}

response = requests.get(url, headers=headers, params=params)

if response.status_code == 200:
    benutzer_daten = response.json()
    print(benutzer_daten)
else:
    print("Es gab ein Problem:", response.status_code)
```

In obigen Python-Beispiel wird die Bibliothek `requests` verwendet, um eine HTTP-Anfrage an eine API zu senden.

- Der String `url` enthält den Endpoint, d.h. den spezifischen Ort auf einem Webserver, an den unsere Anfrage gesendet wird.
- Im Dictionary `headers` speichern wir Anmeldeinformationen, die dem Server mitteilen, dass wir autorisiert sind, auf die gewünschten Daten zuzugreifen.
- Das Dictionary `params` enthält Parameter, die den Zweck unserer Anfrage spezifizieren. Hier bitten wir um Informationen über einen Benutzer mit der ID "12345".

- Die Methode `requests.get()` sendet eine GET-Anfrage an den in `url` angegebenen Server, wobei die zuvor definierten Header und Parameter verwendet werden.
- Schließlich überprüft der Code den Status der Antwort. Wenn der Statuscode 200 zurückgegeben wird, werden die Daten des Benutzers extrahiert und gedruckt. Andernfalls wird eine Fehlermeldung ausgegeben.

1.3 Interaktion mit der Google Cloud API

1.3.1 Umgang mit dem Google Cloud Dienst

1.3.2 Erstellen eines API-Keys in Google Cloud

2 Spracherkennung und -transkription

2.1 Schritte im Speech-to-Text Prozess

2.2 Google Cloud Speech API

3 Sprachübersetzung

3.1 Erkennung und Übersetzung von Texten

3.2 Cloud Translation API

4 Textanalyse

4.1 Klassifizierung von Text in Kategorien

4.2 Entitäten- und Sentimentanalyse

4.3 Natural Language API

5 Sprachsynthese

5.1 Erzeugung synthetischer Sprache

5.2 Natural Language API

Literatur

- [1] Google Cloud. *Language, Speech, Text, & Translation with Google Cloud APIs*. Zugriffsdatum: 08. August 2023. 2023. URL: <https://www.cloudskillsboost.google/quests/179>.