Sprache und Sprachressourcen im digitalen Alltag

Collections
Lexical Resources
Editions
Infrastructure/ Operations

Andreas Witt, Annica Skupch, Thorsten Trippel, Antonina Werthmann
Leibniz-Institut für Deutsche Sprache, Mannheim

Was sind Sprachressourcen?

Elektronisch verfügbare Sprachdaten, die optional mit linguistischen Informationen (z.B. Annotationen) angereichert werden können

- Geschriebene Sprachdaten wie Texte, Lexika, Wortnetze
- Gesprochene Sprache wie Transkripte von Reden, Parlamentsprotokolle und Interviews

Wie werden Sprachressourcen in der KI eingesetzt?

- Regelbasierte Verfahren der Sprachverarbeitung...
  - beruhen auf einem festgelegten Satz von (z. B. syntaktischen oder semantischen) Regeln
  - verwenden Regeln, um Sprache zu analysieren, zu verstehen oder zu generieren
- Statistische Verfahren der Sprachverarbeitung...
  - verwenden maschinelles Lernen und statistische Modelle
  - erlernen Wahrscheinlichkeiten und Muster in Bezug auf Wörter, Phrasen, Grammatikstrukturen und semantische Beziehungen
  - benötigen eine große Menge von sprachlichen Daten

Moderne Systeme in der Sprachverarbeitung nutzen oft eine Kombination aus beiden Ansätzen.

Wo werden Sprachressourcen angewendet?

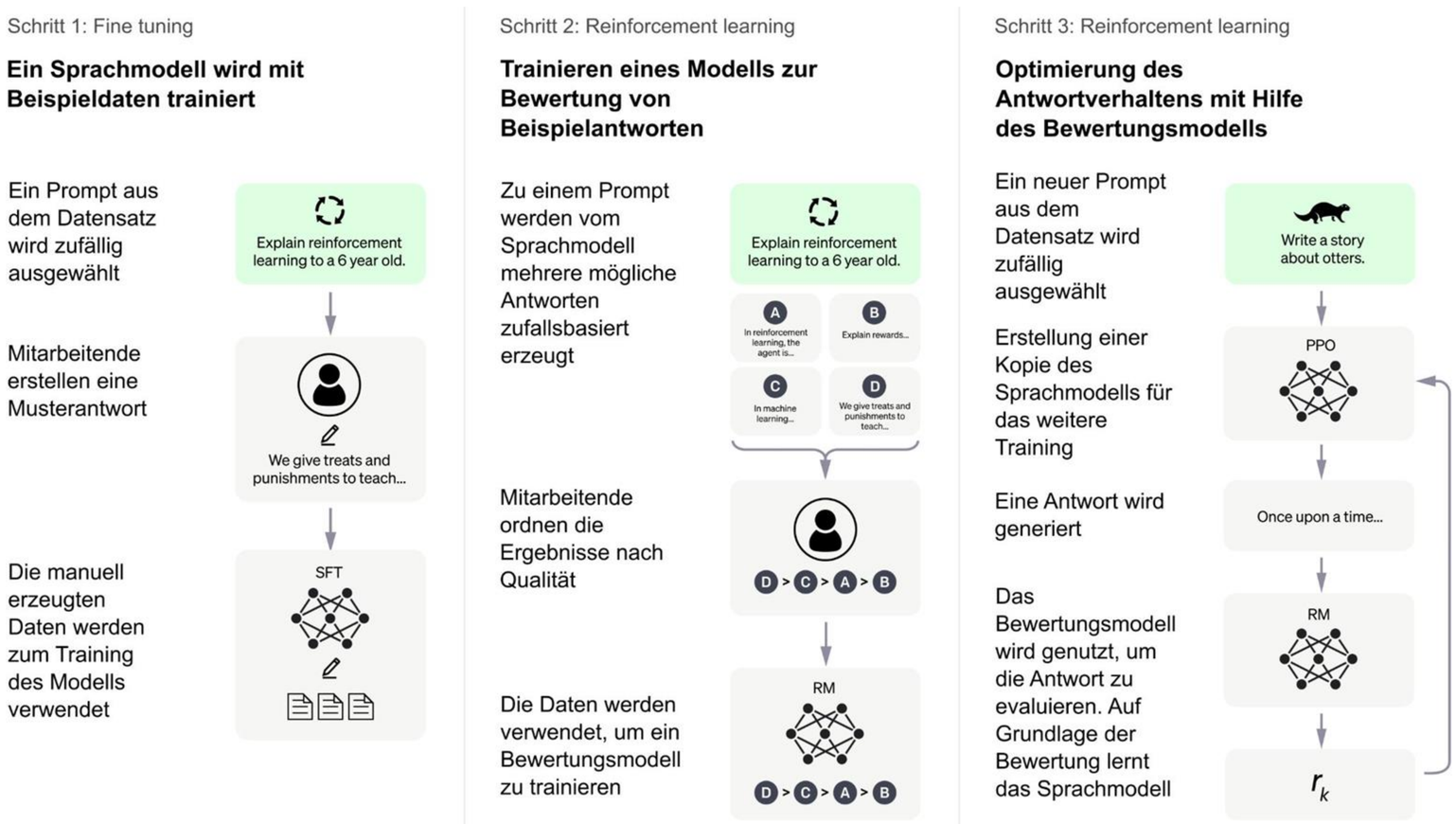
Umfangreiche Sprachdaten bilden eine Grundlage für den Aufbau und das Training funktionaler Sprachmodelle wie ELIZA, GPT oder BERT. Diese Modelle werden in verschiedenen Bereichen angewendet, darunter:

- Forschung, z.B. zur Analyse von Texten oder Extraktion von Informationen.
- Natürliche Sprachverarbeitung (NLP), z.B. bei der maschinellen Übersetzung, Sentiment-Analyse, Text Mining.
- Bildung, z.B. zur Erstellung von Lehrmaterialien und Durchführung automatischer Bewertungen.
- Gesundheitswesen, z.B. zur Analyse und Auswertung von Patientenakten.

Wovon hängt die Qualität der Sprachmodelle ab?

- Quantität der Sprachressourcen, die für das Training angewendet werden
- Qualität der Sprachressourcen, die korrekte Sprache für vielfältige Anwendungsdomänen enthalten
- Repräsentativität der Sprachressourcen, die einen ausgewogenen Querschnitt der Sprache und ihre Vielfalt in verschiedenen Variationen und Kontexten abbilden
- Relevanz der Sprachressourcen, die bestimmte Anwendungsfälle abdecken
- Aktualität der Sprachressourcen, die den gegenwärtigen Gebrauch und die Veränderung der Sprache widerspiegeln
- Diversität der Sprachressourcen, die eine breitere Palette von sprachlichen Mustern, Ausdrücken, Dialekten und spezifischen Nuancen erfassen

Wie funktionieren Sprachmodelle wie ChatGPT?



In Anlehnung an: https://openai.com/blog/chatgpt?ref=assemblyai.com

Teste dein Wissen zu Sprachmodellen und -ressourcen mit ChatGPT!

Beim Scannen des QR-Codes gelangst du zu einem Karteikarten-Quiz, bei dem dir Antworten von ChatGPT gezeigt werden, die auf Fragen zum Thema dieses Posters gegeben wurden. Deine Aufgabe ist es, dir – auf Grundlage der ChatGPT Antworten – die dazugehörigen Fragen zu erschließen.

https://view.genial.ly/64d551c6e915e00012de120a/interactive-content-chatgpt-fragen-und-antworten

Funded by
Deutsche Forschungsgemeinschaft
German Research Foundation
The NFDI consortium Text+ is funded by Deutsche Forschungsgemeinschaft (DFG), project number 460033370.

Text+ is part of
Nationale Forschungsdaten Infrastruktur

https://www.text-plus.org