



Multilingual, Open-text Unified Syntax-independent SEMantics

Ergebnisse in Kürze

Projekt geht neue Wege in der KI, um die „DNS der Sprache“ zu schaffen

Mit neuen Ansätzen des maschinellen Lernens liefert die Forschung sprachunabhängige Repräsentationen von Texten, mit denen KI-Halluzinationen bewältigt werden können.



Halluzinationen der künstlichen Intelligenz (KI) (Erzeugung von Inhalten, die falsch sind und nicht auf realen Daten beruhen) sind durch die Einführung großer Sprachmodelle wie ChatGPT und Bard zu einem Trendthema geworden. Doch über ein EU-finanziertes Forschungsprojekt können Informatikerinnen und Informatiker der perfekten, sprachunabhängigen Verarbeitung natürlicher Sprache näher kommen und dabei den Faktor der Fehlinformation von KI vermeiden.

[MOUSSE](#) , oder Multilingual, Open-text Unified Syntax-independent SEMantics, untersuchte neue Wege, die Fähigkeiten des mehrsprachigen semantischen Parsings zu verbessern, ohne die bedeutende Hürde, Daten für jede einzelne Sprache zu annotieren.

„Große Sprachmodelle wie ChatGPT oder Bard sind zwar leistungsfähig und beeindruckend, haben aber immer noch Schwierigkeiten, die Sicherheit und den gesunden Menschenverstand nachzuahmen, die den Menschen ausmachen.“

MOUSSE schafft die Voraussetzungen für dieses ehrgeizige Ziel. Es bietet ein riesiges Repository an mehrsprachigem Fachwissen, mit dem die Schlussfolgerungen und Ausgaben dieser Modelle begründet und das problematische Phänomen der Halluzination bewältigt werden kann“, erklärt [Roberto Navigli](#), Leiter der [Sapienza Natural Language Processing Group](#) und Projektkoordinator von MOUSSE.

Mehrsprachigkeit als Ressource nutzen

Das von MOUSSE entwickelte umfangreiche Repository wird von Navigli als „die DNS der Sprache“ bezeichnet, da es die Grundlage für die Konstruktion sinnvoller Sätze in vielen Sprachen bildet.

Dies ist dem Hauptergebnis des Projekts zu verdanken: der Fähigkeit, das rechnerische Äquivalent zu den mentalen Darstellungen zu schaffen, die Menschen von Texten erzeugen, jedoch unabhängig von der Sprache. Navigli führt aus: „Der Computer erstellt ein Bild von der Bedeutung eines Satzes und abstrahiert dabei weg von der Sprache und der Oberflächenform, d. h. von den Wörtern, durch die diese Bedeutung ausgedrückt wird.“

Je mehr Sprachen das Team verwendet, um die Semantik auszudrücken, desto mehr kann es die Qualität der erlernten Darstellung ausbauen. Andererseits können, sobald eine Darstellung eines Satzes in einer Sprache erstellt wurde, Sätze mit der gleichen Bedeutung in anderen Sprachen ausgegeben werden. „Das ähnelt der maschinellen Übersetzung, geht aber noch einen Schritt weiter: Es ist ein formaler, strukturierter Nachweis dafür, was die Maschine verstanden hat“, sagt Navigli.

Indem es Mehrsprachigkeit ausnutzt, trägt MOUSSE dazu bei, das Forschungsgebiet der Verarbeitung natürlicher Sprache für alle EU-Sprachen und Hunderte von anderen Sprachen anzugleichen. Das mehrsprachige Repository kann auch beim Spracherwerb nützlich sein, um den Wortschatz zu erweitern und auf eine Weise zu lernen, die mehr auf der Bedeutung als auf einzelnen Wörtern beruht.

Nutzung von KI-Tools für das beste Ergebnis

Die von MOUSSE entwickelten Funktionen werden in vier Hauptschritten erlangt, nämlich Disambiguierung der Wortbedeutung, Entitätsverknüpfung, Markierung der semantischen Funktion und semantisches Parsing. Neben dem Einsatz von Deep Learning wurden die Ergebnisse dadurch erzielt, dass das Modell und dessen Ausgaben interpretierbar und manipulierbar gehalten wurden.

Zusammenfassend konnte das Projekt symbolisches Wissen und neuronale Netze miteinander verbinden und so einen innovativen neuro-symbolischen Ansatz

eingeführen. Laut Navigli wird dadurch das Beste aus beiden Welten vereint: die hohe Leistung und Effektivität der neuronalen Modelle und die Interpretierbarkeit, Manipulierbarkeit und Sprachunabhängigkeit des symbolischen Anteils.

Symbolisches Wissen wird im Wesentlichen durch mehrsprachige Wissensgraphen wie BabelNet bereitgestellt, einem riesigen mehrsprachigen enzyklopädischen Computerwörterbuch, das aus dem ebenfalls von Navigli koordinierten Projekt [MultiJEDI](#) hervorgegangen ist. Die Ergebnisse beider Projekte wurden von seiner erfolgreicher Universitätsausgründung [Babelscape](#) weiterentwickelt und nachhaltig gesichert.

Schlüsselbegriffe

[MOUSSE](#)

[KI](#)

[künstliche Intelligenz](#)

[KI-Halluzination](#)

[großes Sprachmodell](#)

[Verarbeitung natürlicher Sprache](#)

[maschinelles Lernen](#)

[Deep Learning](#)

[Disambiguierung der Wortbedeutung](#)

[Markierung der semantischen Funktion](#)

[semantisches Parsing](#)

[symbolische KI](#)

Entdecken Sie Artikel in demselben Anwendungsbereich



Die Zukunft des Essens durch die „virtuelle Zunge“

5 April 2024





Verbesserte Datenerfassung und -verwaltung in der Bürgerwissenschaft

29 September 2022



Eine Quanteninformationsverarbeitung im größeren Maßstab – von der Theorie zur Realität

29 Mai 2020



Forschung entdeckt Mechanismen für die Steuerung faszinierender Phasenübergänge vom Isolator zum Metall

26 März 2021



Projektinformationen

MOUSSE

ID Finanzhilfevereinbarung: 726487

[Projektwebsite](#)

DOI

[10.3030/726487](https://doi.org/10.3030/726487)

Projekt abgeschlossen

Finanziert unter

EXCELLENT SCIENCE - European Research Council (ERC)

Gesamtkosten

€ 1 497 250,00

EU-Beitrag

€ 1 497 250,00

Koordiniert durch

EK-Unterschriftsdatum

27 Februar 2017

UNIVERSITA DEGLI STUDI DI
ROMA LA SAPIENZA

 Italy

Startdatum

1 Juni 2017

Enddatum

31 Mai 2023

Dieses Projekt findet Erwähnung in ...



5 September 2023



Letzte Aktualisierung: 25 August 2023

Permalink: <https://cordis.europa.eu/article/id/446038-project-breaks-new-grounds-in-ai-to-create-dna-of-language/de>

European Union, 2025