

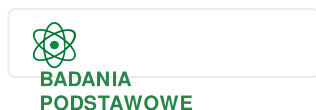


Multilingual, Open-text Unified Syntax-independent SEMantics

Wyniki w skrócie

Przełomowy projekt w dziedzinie sztucznej inteligencji opracuje „DNA języka”

Dzięki nowym podejściom do uczenia maszynowego wykorzystanym przez badaczy po raz pierwszy udało się uzyskać reprezentacje tekstu niezależne od języka, co może pozwolić na rozwiązanie problemu halucynacji modeli sztucznej inteligencji.



Wraz z wprowadzeniem na rynek dużych modeli językowych, takich jak ChatGPT i Bard popularnym tematem stały się halucynacje sztucznej inteligencji – zjawisko generowania nieprawdziwych treści, które nie opierają się na rzeczywistych danych. Dzięki finansowanemu przez UE projektowi badawczemu naukowcy zajmujący się informatyką stawiają kolejny krok na drodze do udoskonalenia technologii przetwarzania języka naturalnego (NLP) niezależnie od

języka, rozwiązując jednocześnie problem fałszywych informacji wytwarzanych przez modele sztucznej inteligencji.

W ramach projektu [MOUSSE](#) , czyli Multilingual, Open-text Unified Syntax-independent SEMantics, badacze skupili się na nowych kierunkach działań mających na celu zwiększenie możliwości w zakresie wielojęzycznego parsowania semantycznego, bez konieczności opisywania danych dla poszczególnych języków.

„Pomimo wyjątkowych możliwości i zapierających dech w piersiach funkcji, duże

modele językowe, takie jak ChatGPT czy Bard, wciąż mają trudności z uzyskaniem zarówno pewności, jak i zdrowego rozsądku, które charakteryzują ludzkie myślenie. Celem projektu MOUSSE jest położenie podwalin pod realizację tego ambitnego celu. Badacze opracowali ogromne repozytorium wiedzy obejmującej wiele języków, która może być wykorzystana do ugruntowania rozumowania i wyników tych modeli oraz radzenia sobie z problematycznym zjawiskiem halucynacji”, stwierdza [Roberto Navigli](#) , kierownik [Grupy ds. Przetwarzania Języka Naturalnego przy Uniwersytecie Sapienza w Rzymie](#)  i koordynator projektu MOUSSE.

Wykorzystywanie wielojęzyczności jako zasobu

Obszerne repozytorium opracowane w ramach projektu MOUSSE jest określane przez Navigliego jako „DNA języka”, ponieważ stanowi podstawę do konstruowania sensownych zdań w wielu językach.

Jest to możliwe dzięki najważniejszemu rezultatowi prac – możliwości tworzenia obliczeniowego odpowiednika mentalnych reprezentacji tekstów przygotowanych przez ludzi, niezależnie od języka. Jak wyjaśnia Navigli: „Komputer tworzy wyobrażenie o znaczeniu zdania, które nie ma związku z językiem i formą samego zdania, czyli słów, za pomocą których jest wyrażane znaczenie”.

Im większej liczby języków używa zespół do wyrażania semantyki, tym bardziej wyraźna staje się jakość wyuczonej reprezentacji. Co więcej, po uzyskaniu reprezentacji zdania w jednym języku, można uzyskać zdania w innych językach w celu wyrażenia tego samego znaczenia. „Proces ten przywodzi na myśl tłumaczenie maszynowe, jednak idzie o krok dalej – zapewnia bowiem formalny, ustrukturyzowany dowód na to, co zrozumiała maszyna”, mówi Navigli.

Wykorzystując wielojęzyczność, projekt MOUSSE przyczynia się do wyrównania szans w zakresie badań nad NLP dla wszystkich języków Unii Europejskiej oraz setek innych. Wielojęzyczne repozytorium może być również przydatne dla osób uczących się języka, pomagając im poprawić słownictwo i uczyć się w sposób oparty w większym stopniu na znaczeniu niż na pojedynczych słowach.

Wykorzystanie narzędzi SI w celu uzyskania najlepszych wyników

Możliwości opracowane w ramach projektu MOUSSE są osiąganе w czterech głównych krokach, które obejmują ujednoznacznienie słów, łączenie jednostek, etykietowanie ról semantycznych i parsowanie semantyczne. Osiągnięcie tych rezultatów było możliwe nie tylko dzięki uczeniu głębokiemu, ale także dzięki zapewnieniu możliwości interpretowania modelu i jego wyników oraz manipulowania nimi.

Podsumowując, projekt był w stanie połączyć wiedzę symboliczną i sieci neuronowe, tworząc w ten sposób innowacyjne podejście neurosymboliczne. Według Navigliego oznacza to, że nowe rozwiązanie pozwala na wykorzystanie najlepszych aspektów obu tych dziedzin – wysoką wydajność i skuteczność modeli opartych na sieciach neuronowych oraz możliwość interpretowania, manipulowania i niezależność językową części symbolicznej.

Zasadniczo wiedza symboliczna opiera się na wielojęzycznych grafach wiedzy, takich jak BabelNet, ogromny wielojęzyczny encyklopedyczny słownik komputerowy, który stanowił rezultat prac w ramach projektu [MultiJEDI](#), również koordynowanego przez Navigliego. Opracowaniem i rozwojem rezultatów obu projektów zajmuje się odnosząca sukcesy spółka spin-off [Babelscape](#), na której czele stanął badacz. Spółka dba także o to, by rezultaty prac były trwałe i zrównoważone.

Słowa kluczowe

[MOUSSE](#)

[SI](#)

[sztuczna inteligencja](#)

[halucynacje SI](#)

[duży model językowy](#)

[przetwarzanie języka naturalnego](#)

[NLP](#)

[uczenie maszynowe](#)

[głębokie uczenie](#)

[ujednoznacznienie słów](#)

[etykietowanie ról semantycznych](#)

[parsowanie semantyczne](#)

[symboliczna sztuczna inteligencja](#)

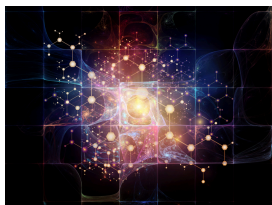
Znajdź inne artykuły w tej samej dziedzinie zastosowania



Przetwarzanie kwantowe staje się rzeczywistością

29 Maja 2020





Badania ujawniają mechanizmy rządzące niesamowitymi przemianami fazowymi izolator-metal

26 Marca 2021



Uczenie sieci neuronowych jak płynąć z prądem

26 Lutego 2021



Jak nasz mózg wykorzystuje wzrok do planowania pewnego chwytania przedmiotów

18 Grudnia 2020



Informacje na temat projektu

MOUSSE

Identyfikator umowy o grant: 726487

[Strona internetowa projektu](#)

DOI

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

Projekt został zamknięty

Finansowanie w ramach

EXCELLENT SCIENCE - European Research Council (ERC)

Koszt całkowity

€ 1 497 250,00

Wkład UE

€ 1 497 250,00

Koordynowany przez

Data podpisania przez KE
27 Lutego 2017

UNIVERSITA DEGLI STUDI DI
ROMA LA SAPIENZA

 Italy

Data rozpoczęcia
1 Czerwca 2017

Data zakończenia
31 Maja 2023

Ten projekt został przedstawiony w...



5 Września 2023



Ostatnia aktualizacja: 25 Sierpnia 2023

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

European Union, 2025