



# VYUŽITIE NEURÓNOVÝCH SIETÍV DETEKCII PLAGIÁTOV

ORYDOROHA BOHDAN, ROIKO OLEKSII, TSIMBOTA VLADYSLAV A ZELENSKA ZLATA

# ČO JE PLAGIÁTORSTVO

- Plagiátorstvo je nečestným činom, ktorým sa používa cudzia práca alebo myšlienky bez uznania pôvodného zdroja. Toto porušenie môže mať závažné následky v akademickom aj profesionálnom prostredí. Plagiátorstvo môže obsahovať kopírovanie textu, obrázkov, hudby alebo iných intelektuálnych vlastníctiev bez súhlasu a predstavovanie ich ako vlastné dielo.



# PLAGIÁTORSTVO A AUTORSKÉ PRÁVA

- Plagiátorstvo sa prelína s autorským právom a literárnymi vlastnickými právami. Autorské právo chráni práva tvorcov a vlastníkov originálnych diel, ako sú knihy, články, hudba a softvér. Keď niekto plagiuje, v podstate porušuje tieto práva použitím diela bez súhlasu a bez riadneho uvedenia zdroja.





## PRÍKLADY

- Plagiátorstvo môže byť napríklad skopírovaním odseku z publikovaného článku do eseje alebo začlenením častí niekoho iného románu do vlastného diela bez uvedenia zdroja. Takéto konanie je nezákonné a môže mať právne dôsledky. Aj skvelé mysle, ako bola Nikola Tesla, môžu byť zatienené neprávom ostatnými vynálezcami a obchodníkmi. Našťastie, využívanie technológií, ako sú neurónové siete, môže pomôcť pri odhaľovaní a prevencii plagiátorstva.



## VÝHODY POUŽITIA NEURÓNOVÝCH SIETÍ PRI DETEKCII PLAGIÁTOV

- Neurónové siete sú veľmi presné pri identifikácii vzorov a korelácií v rámci veľkých súborov údajov.
- Sémantické porozumenie umožňuje detekciu aj parafrázovaného alebo prepísaného obsahu.
- Adaptabilita umožňuje prispôsobenie sa meniacim sa trendom v jazyku.
- Škálovateľnosť umožňuje efektívne spracovanie rozsiahlych súborov údajov.



# ZHRNUTIE VÝHOD A PERSPEKTÍVY

- Použitie neurónových sietí je podporované rôznymi štúdiami, ktoré dokazujú ich vysokú presnosť, sémantické porozumenie, adaptabilitu a škálovateľnosť.
- Ich schopnosť adaptovať sa na nové vzory a štýly textu umožňuje rýchlu a automatickú detekciu prípadov plagiátorstva v reálnom čase.
- S pokrokom v oblasti hlbokého učenia a analýzy textu sa očakáva, že neurónové siete sa budú ďalej zlepšovať a zvyšovať presnosť a účinnosť detekcie plagiátorstva.





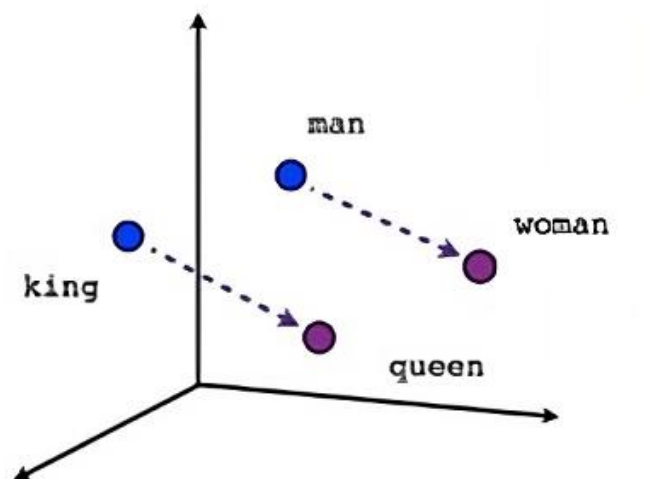
# PREHĽAD APLIKÁCIÍ NEURÓNOVÝCH SIETÍ V DETEKCII PLAGIÁTORSTVA

- Neurónové siete sa trénujú na veľkých množstvách textových dát pre identifikáciu podobností medzi rôznymi textami
- Dve hlavné formy plagiátu: parafrázovanie a patchwork
- Neurónové siete analyzujú text tým, že ho rozdeľujú na menšie časti a porovnávajú s databázou existujúcich textov
- Rýchly a automatický proces detekcie plagiátorstva

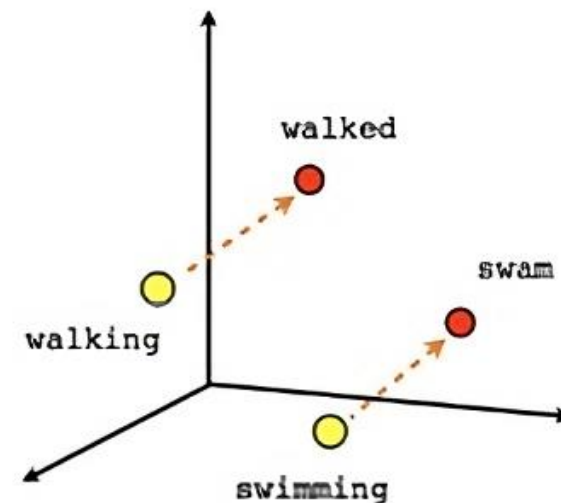


## KĽÚČOVÉ KONCEPTY A TECHNIKY PRE NEURÓNOVÉ SIETEV DETEKCII PLAGIÁTORSTVA

- Vektorová reprezentácia textu – embedding
- Word2Vec, GloVe, BERT
- Rekurentné neurónové siete (RNN)
  - Práca s časovými radmi alebo sekvenciami, kontext a závislosti
  - LSTM a GRU vrstvy pre dlhodobé závislosti



Male-Female



Verb tense



# POKROČILÉ TECHNIKY PRE NEURÓNOVÉ SIETE V DETEKCII PLAGIÁTORSTVA

## I. Siamské neurónové siete

- Porovnávanie dvoch alebo viacerých vstupov na identifikáciu podobností alebo rozdielov

## II. Konvolučné neurónové siete (CNN)

- Extrahovanie miestnych črt z textu, napr. n-gramy alebo časté slovné spojenia
- Kombinácia s RNN alebo siamskými neurónovými sieťami pre presnejšiu analýzu textových dát

## III. Manuálne preverenie identifikovaných častí textu na zistenie skutočného plagiátorstva

- Skúmanie kontextu a účelu použitia častí textu
- Posúdenie, či ide o skutočné plagiátorstvo alebo iný dôvod podobnosti
- Následné kroky v prípade potvrdenia plagiátorstva





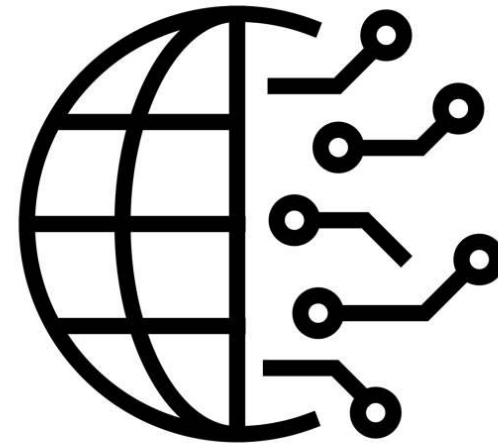
# KĽÚČOVÉ KRITÉRIÁ POROVNÁVANIA A TRANSFORMÁCIA TEXTU

## Vektorová reprezentácia textu

- Transformácia textu na zoznam vektorov
- Udržiavanie sémantických a syntaktických vlastností pomocou algoritmov hlbokého učenia

## Úroveň spracovania textu

- Určovanie, či je text spracovaný na úrovni slov alebo viet
- Transformácia textu na zoznam viet ako najvhodnejšia reprezentácia



# METÓDY VÝPOČTU PODOBNOSTI A ICH OBMEDZENIA

- Architektúry CNN a RNN
  - Väčšina prístupov sa spolieha na vektorovú reprezentáciu na úrovni slov
  - Obmedzenie v zisťovaní podobnosti medzi celými vetami alebo textami
- Kosínusová podobnosť
  - Nespolahlivé výsledky v prípade sémanticky nepodobných dokumentov s rovnakými slovami alebo vetami
- Potreba prístupu zohľadňujúceho sémantiku
  - Text reprezentovaný ako zoznam viet transformovaný na zoznam vektorov
  - Použitie algoritmov, ako je RNN, na zisťovanie podobnosti



# APLIKÁCIE NEURÓNOVÝCH SIETÍ V DETEKCII PLAGIÁTORSTVA

- I. Rôzne odvetvia a kontexty
  - Akademické prostredie, vydavateľský priemysel, právne odvetvie
- II. Nástroje a platformy pre detekciu plagiátorstva
  - Turnitin, iThenticate, PlagScan
  - Nahrať dokumenty a analyzovať ich na potenciálne prípady plagiátorstva



# ZHODNOTENIE A ZÁVER: NEURÓNOVÉ SIETE V BOJI PROTI PLAGIÁTORSTVU

Rastúci problém  
plagiátorstva

- Technológie ako ChatGPT a ďalšie zvyšujú výskyt plagiátorstva

Použitie RNN a  
Word2Vec

- Dosiahnutá presnosť detekcie plagiátov: 61%

Neurónové siete ako  
účinný a efektívny spôsob  
boja proti plagiátorstvu

- Rastúca popularita v mnohých odvetviach

Dôležitosť oboznámenia  
sa s technológiou

- Začlenenie neurónových sietí do činností na zabezpečenie integrity a autenticity obsahu

