

# Využitie Natural Language Processing pre lepšie učenie jazyka\*

Branislav Hozza

Slovenská technická univerzita v Bratislave  
Fakulta informatiky a informačných technológií  
xhozza@stuba.sk

8. október 2020

## Abstrakt

V tejto práci sa zameriavame na zlepšenie učenia jazyka pomocou NLP [3]. Výhody NLP majú bohaté využitie v učení ako napríklad prístup k obrovskému zdroju textov. NLP sa dnes využíva v mnohých odvetviach či už ako nástroje na rozpoznávanie reči alebo pomocník na opravu gramatických chýb. V edukačnom systéme sa dá hlavne využiť na výučbu jazykov. NLP dokáže vyhodnocovať komplexnosť textu a kontrolovať gramatiku či plagiátorstvo. Tento dokument bude zameraný hlavne na možné využitia a výzvy, ktorým musíme čeliť pri využití NLP technológie. Zameriam sa na to ako NLP učí o jazykoch a ako učí samotný jazyk.

## 1 Úvod

NLP<sup>1</sup> sa v dnešnej dobe používa takmer v každom odvetví, či už v zdravotníctve, v informatike, kontrola textu, dátová analýza, atď. S príchodom nových technológií ako sú napríklad Big Data, Machine learning, MOOC<sup>2</sup> sa otvorili mnoho možností, ako využiť NLP v praxi. V produkcii je už aktuálne mnoho aplikácií, ktoré túto technológiu využívajú pri rôznych kurzoch. V edukačnom systéme si môžeme ukázať v zozname 1.

Táto technológia má veľký potenciál vo využití pri učení anglického jazyka alebo STEM<sup>3</sup> predmetoch. Pri tejto forme štúdia vieme NLP využiť na kontrolu gramatiky pri písaní esejí, učenie slovíčok alebo správnej výslovnosti. Ako najväčšiu výhodu využitia tejto technológie v praxi vidíme hodnotenie testov so skalárne veľkým počtom testovaných študentov. Okrem korektnosti testov sme schopní taktiež kontrolovať plagiátorstvo textu.

Ako môžeme vidieť v obraze Č.1, výskum v aplikovaní NLP v edukácii spočíva v opakujúcom sa cykle. Technologické inovácie sú inšpirované socialnými potrebami. Technologický výskum je najprv informovaný a následne prispieva do edukačnej teórie a poskytuje dáta.

---

\*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2020/21, vedenie: J. Sitarčík

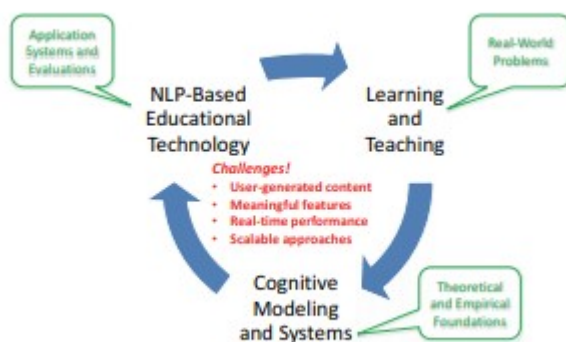
<sup>1</sup>Natural Language Processing - spracovanie prirodzeného jazyka

<sup>2</sup>Massive open online course - veľké online kurzy

<sup>3</sup>Science, technology, engineering and mathematics - veda, technológia, inžinierstvo a matematika

- Vyučovanie lingvistických predmetov. – napr., čítanie, písanie, rozprávanie
- Používanie NLP v potrebách študentov alebo učiteľov – napr., knihy, učebné materiály, softvér
- Učenie matematiky alebo fyziky – napr., vytváranie slovných úloh, generovanie príkladov

Zoznam 1



Obr. 1: Cyklus technologickej inovácie

## 2 Natural Language Processing

Je to druh umelej inteligencie, zameranej na prácu s textom a obrázkami. Táto technológia je spojením lingvistiky a informatiky, pričom vzniká snaha aby stroj porozumel prirodzenej reči človeka. NLP má za sebou 70 rokov vývoja a prvé zmienky sú z roku 1950 [1].

Už v roku 1950, Alan Turing publikoval článok "Computing Machinery and Intelligence" [4] ktorý priniesol tzv. Turingov test ako kritérium inteligencie, pre automaticky generovanú prirodzenú reč. V tej dobe sa NLP nerozlišovalo od umelej inteligencie. Okolo roku 2010 sa začali rozširovať metódy machine learningu ako deep learning alebo representation learning aj v NLP, pretože bolo dokázané že tieto techniky sú veľmi účinné.

### 3 Učenie o jazyku

Jedným z najstarších a napriek tomu najaktívnejším využití NLP je jeho využitie pri učení jazyka. Tento proces zvyčajne pozostáva vo vyhodnotení zručnosti študenta v písaní testov, v čítaní alebo rozprávaní v danom jazyku. Syntaktická analýza sa používa na detekciu a potencialne opravenie chybného použitia predložiek pri skupinách ako sú ESL<sup>4</sup> alebo hluchý študenti.

### 4 Učenie jazyka

Narozdiel od využitia ako analýzy (viď. predošlá časť), jazyk môže byť využitý aj ako výučbová metóda. V článku z roku 2011 [2] bolo dokázané, že skupina detí ktoré sa učia pomocou živého učiteľa, pochopili látku lepšie ako žiaci, ktorí sa učili pomocou PC. Hlavným rozdielom medzi ľudským učiteľom a PC učiteľom je, že iba človek vie používať prirodzené vety.

V posledných rokoch sa systémy založené na dialógu snažia výkonovo čo najviac priblížiť prirodzenej ľudskej reči. Uvažovalo sa aj nad tým, že učenie by sa malo odohrávať vo viac socialno realistickom prostredí aby sa PC učiteľ vedel viac adaptovať študentom.

### 5 Spracovanie jazyka

Okrem úlohy NLP ako prostriedok na spracovanie textu a hovorenej reči pri vyhodnocovaní testov, alebo ako médium medzi učivom a výkladom, má NLP aj veľké využitie pri spracovaní textu pre vedcov, učiteľov alebo vývojárov týchto automatizovaných systémov. Využíva sa obrovské množstvo textov a zvukových stôp, ktoré sú dostupné v elektronickej podobe na sociálnych sieťach, fórach, blogoch alebo stránkach ako napríklad Wikipédia. Jedným z najväčších zdrojov je tzv. Linguistic Data Consortium<sup>5</sup>

### 6 Záver

Tento článok bol zameraný na výskum vo využití NLP pri výučbe anglického jazyka. Tento článok bol motivovaný potrebami učiteľov a ľuďmi ktorí sa radi učia jazyky. Aj napriek tomu že využitie NLP v

### Literatúra

- [1] [https://en.wikipedia.org/wiki/Natural\\_language\\_processing#History](https://en.wikipedia.org/wiki/Natural_language_processing#History).
- [2] Chi M.; VanLehn K.; Litman D.; and Jordan. An evaluation of pedagogical tutorial tactics for a natural language tutoring system: A reinforcement learning approach. *International Journal of Artificial Intelligence in Education*, 21(1-2)(83–11), 2011.

<sup>4</sup>English as second language - ľudia pre ktorých angličtina nie je primárny jazyk.

<sup>5</sup>Linguistic Data Consortium (<http://www ldc.upenn.edu/>)

- [3] Diane J Litman. Natural language processing for enhancing teaching and learning. In *AAAI*, pages 4170–4176, 2016.
- [4] Alan M Turing. Computing machinery and intelligence. In *Parsing the turing test*, pages 23–65. Springer, 2009.