



# **Strojový překlad využívající umělou inteligenci**

Ingrid Habermannová



## **Strojový překlad využívající umělou inteligenci**

Už od začátku 50. let se rozvíjí obor strojového překladu. Jak se za tu dobu vyvinul, jaké jsou jeho podoby?

Umělá inteligence může v mnohém pomoci. Ukážeme si, jak vytrénovat překladač.



# Strojový překlad využívající umělou inteligenci

Druhy strojového překladu:

- **„pravidlový“**, založený na znalosti jazyka
- **statistický**, využívající metod strojového učení
- **hybridní**, kombinace obou přístupů



# Znalostní (transferový) strojový překlad

Tradiční přístup.

**Přímý** překlad využívá pouze slovník.

Vyšší úrovně už provádějí:

- morfologickou a **syntaktickou** analýzu, transfer a **generování** gramatikou cílového jazyka
- pomocnou **sémantickou** analýzu
- či dokonce překládají prostřednictvím na přirozeném jazyku nezávislé reprezentace, **interlinguy**



## Statistický strojový překlad

Vyžaduje rozsáhlé korpusy: dvojjazyčný **překladový** a korpus **cílového** jazyka.

Na základě prvního korpusu se pro vstupní text hledá **nejpravděpodobnější** (hrubý) překlad a ten se podle **vyhledá** podle druhého korpusu.

Na obou se nejdříve SMT musí **natrénovat**, jak je možné tomu pomoci?



## Intelligentní agent, který se naučil překládat

Byl navržen systém, který se umí naučit překládat s využitím dvou **jednojazyčných** korpusů obsahujících stejné texty.

Je schopen využít nová data a **zlepšovat** úspěšnost svého překladového stroje.

Jak to dělá?



# Intelligentní agent, který se naučil překládat - ukázka algoritmu

---

## Algorithm VI.1 Agent Algorithm

---

**Require:**  $D_{L_1} = \emptyset$   $D_{L_2} = \emptyset$  a SMT trained with a parallel corpus  $T_0$ ,  $\alpha$ ,  $f$ .

- 1: Create an Inverted Index ( $I$ ) using  $D_{L_2}$
  - 2: **while**  $D_{L_1} = \emptyset$  **do**
  - 3:    $S_{L_1} \leftarrow k$  randomly selected sentences from  $D_{L_1}$
  - 4:   **for**  $s_{L_1} \in S_{L_1}$  **do**
  - 5:      $mt \leftarrow \text{SMT}(s_{L_1})$
  - 6:     Find a set of possible similar sentences  $S_{L_2}$  of  $mt$  querying  $I$
  - 7:     **if**  $f(S_{L_2}, \alpha) == \text{True}$  **then**
  - 8:       Add  $(s_{L_1}, s^*)$  to  $M$ , where  $s^*$  is the best element of  $S_{L_2}$  according to  $f$
  - 9:     **end if**
  - 10:   **end for**
  - 11:    $T_i \leftarrow M$  and the training data  $T_{i-1}$  at previous iteration
  - 12:   Train the SMT using  $T_i$
  - 13:   Test the performance of the new model
  - 14:   Remove  $M$  from  $D_{L_1}$
  - 15: **end while**
-



## Intelligentní agent, který se naučil překládat

0. Nejdříve je jeho překladač natrénován na **zarovnaných** větách (s označenými dvojicemi odpovídajících si vět).
1. Vybere se část vět ve zdrojovém jazyce (ZJ), a přeloží se (P).
2. Slova každé přeložené věty se vyhledají v inverzním indexu vět v cílovém jazyce (CJ) a zjistí se pro ně míra shody (P a CJ).
3. Spočítá se i kvalita překladu (skóre Bleu, porovnáním P a CJ).
4. Nejvýše hodnocená věta P je prohlášena za překlad ze ZJ do CJ a jejich dvojice se přidá do příštího trénovacího korpusu.





## Intelligentní agent, který se naučil překládat

5. Překladač se s nově přidanými větami znovu natrénuje. Vzorek s nimi se už příště nepoužije.
6. S každým opakováním bodů 1–5 se úspěšnost překladu zvyšuje. Cyklus končí, když dojdou věty.



## Zdroje

*History of machine translation* [online]. Dostupné z:  
[http://en.wikipedia.org/wiki/History\\_of\\_machine\\_translation](http://en.wikipedia.org/wiki/History_of_machine_translation).

Vít Baisa. *Studijní materiál PLIN019 Strojový překlad*. Brno:  
Masarykovy univerzita.

Marco Turchi, Tijl De Bie, Nello Cristianini. *An Intelligent Agent that Autonomously Learns how to Translate* [online]. Bristol:  
University of Bristol (UK). Dostupné z:  
[https://patterns.enm.bris.ac.uk/files/Turchi\\_DeBie\\_Cristianini\\_wi\\_iat\\_0.pdf](https://patterns.enm.bris.ac.uk/files/Turchi_DeBie_Cristianini_wi_iat_0.pdf).

Děkuji za pozornost :-)



Karel Mačalik @maca89 cnews.cz