

Z Wikipedie, otevřené encyklopedie

Strojový překlad (angl. *machine translation*) je proces automatického [překladu](#) z jednoho [přirozeného jazyka](#) do jiného pomocí počítače. Strojový překlad byl poprvé použit v 50. letech v USA. Úkol v oblasti [zpracování přirozeného jazyka](#) a [umělé inteligence](#) stále však nebyl plně vyřešen. Poslední technologií jsou neuronové strojové překlady (neuron machine translation, NMT), které se blíží k tzv. lidské paritě neboli shodě s tím, jak by bilingvní osoba přeložila daný text a zda se přeložené shoduje s NMT.^[1] Výsledky jsou již dostatečně kvalitní pro použití v mnoha oblastech, kde pomáhají lidským překladatelům.

Historie strojového překladu

Pokusy o strojový překlad cizojazyčných textů začaly krátce po [2. světové válce](#). Očekávalo se, že pro [počítače](#), jejichž první prototypy byly v té době uvedeny do provozu, nebude tento úkol nijak složitý. Brzy se však ukázalo, že tento předpoklad byl zcela mylný.

[První systém pro strojový překlad](#) byl veřejnosti představen [7. ledna 1954](#) v ústředí firmy [IBM](#). Událost přitáhla značnou pozornost médií i široké veřejnosti. Systém samotný byl ale z dnešního pohledu velmi jednoduchý, jeho slovník obsahoval pouze 250 slov a při předvádění překládal pouze 49 zvlášť vybraných vět z [ruštiny](#) do [angličtiny](#). I přes tuto jednoduchost ale systém vzbudil dojem, že praktické nasazení strojového překladu už není daleko, což pomohlo získat dostatečné finanční prostředky na další výzkum v tomto oboru.

Hlavní motivací pro vývoj dalších podobných systémů byla eskalující [studená válka](#). [USA](#) se začaly obávat následku tehdejšího [Sovětského svazu](#) ve vývoji raketové i jiné vojenské techniky a snažily se pečlivě monitorovat pokrok sovětské vědy. První systémy pro strojový překlad byly proto používány především pro překládání ruskojazyčných vědeckých a odborných publikací do angličtiny. Kvalita překladu však byla velmi nízká a takto vzniklé texty se používaly pouze pro získání základní orientace v obsahu dokumentu a rozhodování, zda stojí za to nechat dokument přeložit profesionálním překladatelem.

V průběhu dalších desetiletí bylo do vývoje v oblasti strojového překladu investováno značné úsilí jak v komerční, tak i v akademické sféře.

Dosahované pokroky byly ale po celou dobu velmi skromné a ani dnešní nejmodernější systémy nedokáží překládat na úrovni, která by byla srovnatelná s člověkem.

Metody strojového překladu

Pravidlový strojový překlad

Nejstarším přístupem ke strojovému překladu je pravidlový přístup, obvykle označovaný jako RBMT (Rule-based Machine Translation). Pravidlový překladový systém obvykle sestává ze slovníku (který může být vytvořen ručně či poloautomaticky) a sady překladových pravidel. Tato pravidla analyzují zdrojovou větu a za pomoci slovníku ji transformují do cílového jazyka. Vývoj pravidlového systému je časově náročný (obvykle se hovoří o desítkách let potřebných pro vytvoření reálně použitelného systému) a vyžaduje práci skupiny jazykových a počítačových expertů.

Pravidlových překladových systémů, které byly nasazeny do praxe, je velice málo a existují jen pro několik málo jazykových párů, zejména angličtinu s francouzštinou (částečně díky [Kanadě](#) a jejím jazykovým zákonům, vyžadujícím aby mnohé dokumenty byly k dispozici v anglické i francouzské verzi). Pravidlové systémy se přesto osvědčily pro překlady textů z velmi omezené jazykové domény, v kterémžto případě je možné slovníkem a sadou pravidel téměř plně postihnout texty, které je nutné překládat. V těchto situacích pravidlový překlad často dosahuje lepší kvality než obecně úspěšnější statistický překlad. Příkladem takového systému může být METEO, kanadský systém pro překlad [předpovědi počasí](#).

Statistický strojový překlad

Zvyšování výpočetní síly počítačů umožnilo v 90. letech 20. století masivní nástup statistických metod do počítačové lingvistiky, což v mnoha jejích oblastech způsobilo doslova revoluci. Nejinak tomu bylo u strojového překladu. Najednou již nebylo potřeba jazykových expertů desítek let vývoje. Ukázalo se, že dostatečné množství paralelních jazykových dat a statistické modelování dokáží v krátkém čase a s nevelkými náklady překonat mnohé

stávající pravidlové systémy; stejně tak stvoření překladového systému pro nový jazykový pár se prakticky stalo pouze otázkou získání dostatečného množství dat.

Základem [statistického překladového systému](#) (SMT, Statistical Machine Translation) je dvojjazyčný (paralelní) korpus – sada dokumentů ve zdrojovém jazyce a jejich lidských překladů do cílového jazyka (nebo obráceně), typicky automaticky sesbíraných z internetu a dalších volně dostupných zdrojů. Z těchto dat si překladový systém extrahuje překladovou tabulku, tj. slovník obohacený o pravděpodobnosti (četnosti) jednotlivých možných překladů. Nejobvyklejším přístupem je frázový překlad, kdy překladová tabulka (zde nazývaná frázová tabulka) obsahuje nejen překlady jednotlivých slov, ale zejména krátkých frází – je tomu tak mimo jiné proto, že slova zdrojového a cílového jazyka si obvykle neodpovídají 1:1, a překlad celých víceslovných frází proto obvykle vede k lepším výsledkům.

Druhou zásadní částí statistického překladače je jazykový model pro cílový jazyk, který má za úkol z možných překladů jednotlivých frází vybrat takovou kombinaci, která vede ke koherentní a smysluplné cílové větě. Jazykový model se vytváří na základě jednojazyčného (monolingualního) korpusu cílového jazyka. Pokud takový není k dispozici, lze použít i cílovou stranu paralelního korpusu, ale v typickém případě je jednojazyčných dat k dispozici řádově více než dvojjazyčných.

Komerčně nejúspěšnějším systémem se stal [Google Translate](#), který během několika let nabídl překlad mezi desítkami světových jazyků, v kvalitě obvykle postačující pro základní porozumění textu, v mnoha případech dokonce již umožňující správné porozumění většině obsahu přeloženého textu. Systém je pro běžné uživatele [dostupný zdarma online](#), pro profesionální použití je možné používat placené API.

V akademické sféře se prosazuje open-source překladový systém Moses, který je [dostupný zdarma na internetu](#).

Hybridní strojový překlad

Hybridní překlad se snaží kombinovat výhody pravidlových a statistických systémů. Zásadní výhodou statistického přístupu je výrazně větší pokrytí než u pravidlového. Naopak pravidlové systémy obvykle dobře postihují

gramatické jevy, které se obvykle řídí danými pravidly, zatímco statistické systémy se gramatická pravidla snaží (mnohdy neúspěšně) odhadovat z dat, bez lingvistických znalostí.

Neuronový strojový překlad

Neuronový strojový překlad, založený na hlubokém učení, dosáhl v posledních letech rychlého pokroku.

Překladače založené na práci s umělou neuronovou sítí jsou běžně dostupné od roku 2017. Google nasadil v tomto roce neuronovou síť ve svém Překladači [2] a vznikl též konkurenční projekt DeepL, o němž se má za to, že od roku 2022 obvykle poskytuje nejlepší výsledky strojového překladu, obvykle stále vyžadují následnou úpravu člověkem. [3][4][5] V porovnání s Google Překladačem též dosahuje lepších výsledků. [6]

Strojový překlad v Česku

Strojový překlad má v České republice dlouhou a bohatou tradici, vznikla zde celá řada úspěšných akademických i komerčních systémů.

Akademické systémy

Centrem akademického výzkumu je už po několik desetiletí Ústav formální a aplikované lingvistiky při Matematicko-fyzikální fakultě Univerzity Karlovy. V průběhu sedmdesátých a osmdesátých let zde byla pod vedením prof. Petra Sgalla vytvořena rozsáhlá teorie pro formální popis přirozeného jazyka pomocí závislostní syntaxe nazvaná funkčně-generativní popis. Na těchto teoretických základech pak na přelomu sedmdesátých a osmdesátých let minulého století vznikl pod vedením Zdeňka Kirschnera experimentální anglicko-český strojový překladač APAČ a o několik let později podobný systém pro překlad mezi češtinou a ruštinou nazvaný RUSLAN.

Komerční systémy

Komerční systémy pro strojový překlad přišly na český trh krátce po nástupu osobních počítačů na začátku devadesátých let a poměrně rychle si mezi uživateli získaly značnou popularitu.

PC Translator

PC Translator byl první a komerčně nejúspěšnější softwarový nástroj pro překlad cizojazyčných textů na českém trhu. Během 30 let distribuce (od r. 1988) měl více než 40 tisíc legálních uživatelů. Pomáhal uživateli při překladu tím, že rozdělil větu na jednotlivá slova a skupiny slov a pro každé vyhledal ve slovníku několik nevhodnějších překladů. Výsledek zobrazil v přehledné tabulce, podle níž pak uživatel sám sestavoval finální text. Novější verze byly doplněny i o skloňování a časování, anglická verze navíc o větný parser. Překladač umožňoval překlad ze šesti jazyků (angličtina, němčina, ruština, francouzština, italština a španělština) do češtiny a slovenštiny. Autorem programu byl Ing. Rostislav Janča. Program se na českém trhu velmi rychle etabloval a v České republice stále patří mezi nejpoužívanější překladatelské nástroje. Výrobcem a distributorem programu je od roku 1997 firma Langsoft s.r.o.

Transen

Transen byl první překladač na českém trhu, který dokázal v přeloženém textu sám skloňovat nebo časovat česká slova. K tomuto účelu obsahoval podrobně rozpracovanou databázi českých ohýbacích vzorů a sadu základních pravidel české syntaxe. Anglická syntaxe v něm byla popsána pomocí větných šablon sestavených podle knihy "A Guide to Patterns and Usage in English" významného britského lingvisty A. S. Hornbyho. Klíčové součásti překladače vyvinul bratislavský programátor Jozef Baruník, původně pro překlad z angličtiny do slovenštiny. Systém byl na českém trhu nabízen v první polovině devadesátých let, ale nedosáhl většího rozšíření.

Skik

Překladač Skik byl na český trh uveden v polovině devadesátých let. Dokázal provádět základní analýzu syntaxe anglické věty a uměl i skloňovat nebo časovat česká slova v přeloženém textu. Autorem překladače byl Vladimír Smetáček. Program se na českém trhu prodával přibližně deset let, poté jeho autor své komerční aktivity ukončil a vyvinuté know-how nabídl české akademické komunitě.

Eurotran

Překladač Eurotran přišel na český trh na konci devadesátých let a byl prvním překladačem, který byl zaměřený na překlad internetových stránek. Autorem byl Milan Šustek. Program používal překladový model založený na paralelní bezkontextové gramatice, později přidal statistický bigramový model pro vyhlazování českého textu. Díky nasazení inovativních technik byl hned od počátku nástrojem, který poskytoval v češtině nejvyšší kvalitu překladu.^[7]

Související články

- [Babelfish](#)
- [DLT](#) - projekt *Distributed Language Translation*
- [Interlingvální strojový překlad](#)
- [Transferový strojový překlad](#)
- [Posteditace](#)
- [Statistický strojový překlad](#)
- [Wikipedie:Strojový překlad Wikipedie](#) – Projekty automatizovaného překladu Wikipedie

Reference

1. ↑ Wayback Machine. web.archive.org [online]. [cit. 2023-08-17]. [Dostupné v archivu](#) pořízeném z [originálu](#) dne 2023-08-17.
2. ↑ Google skokově vylepšil překlad. Umělá inteligence se zakousla do češtiny. *iDNES.cz* [online]. 2017-04-19 [cit. 2023-08-17]. [Dostupné online](#).
3. ↑ KATSNELSON, Alla. Poor English skills? New AIs help researchers to write better. *Nature*. 2022-08-29, roč. 609, čís. 7925, s. 208–209. [Dostupné online](#) [cit. 2023-08-17]. [doi:10.1038/d41586-022-02767-9](https://doi.org/10.1038/d41586-022-02767-9). (anglicky)
4. ↑ KORAB, Petr. DeepL: An Exceptionally Magnificent Language Translator. *Medium* [online]. 2022-02-18 [cit. 2023-08-17]. [Dostupné online](#). (anglicky)
5. ↑ DeepL outperforms Google Translate – DW – 12/05/2018. *dw.com* [online]. [cit. 2023-08-17]. [Dostupné online](#). (anglicky)

6. [↑ COLIN. DeepL Translator Review: Is It Better Than Google Translate? \[online\]. 2022-05-18 \[cit. 2023-08-17\]. Dostupné online.](#) (anglicky)
7. [↑ MF DNES: Stroj živého překladače hněd tak nahradí](#), test překladačů PC Translator, Skik a Eurotran.

Externí odkazy

- Obrázky, zvuky či videa k tématu [strojový překlad](#) na Wikimedia Commons
- [Strojový překlad](#) v [České terminologické databázi knihovnictví a informační vědy \(TDKIV\)](#)
- [Online překladač](#) pomocí GNU/FDL slovníku (slovník je [svobodný software](#)). Nejedná se přímo o překlad, ale o zobrazení významu v jiném jazyce v bublinkové nápovedě.
- [Interaktivní ukázka strojového překladu](#) na Ústavu formální a aplikované lingvistiky [MFF UK](#).
- [Strojový překlad Český - Angličtina](#)

Citováno z „https://cs.wikipedia.org/w/index.php?title=Strojový_překlad&oldid=25305562“

[Kategorie:](#)

- [Strojový překlad](#)
- [Umělá inteligence](#)
- [Počítačová lingvistika](#)

Skryté kategorie:

- [Monitoring:Články s odkazem na TDKIV](#)
- [Monitoring:Články s identifikátorem NKC](#)
- [Monitoring:Články s identifikátorem PSH](#)
- [Monitoring:Články s identifikátorem TDKIV](#)
- [Monitoring:Články s identifikátorem BNF](#)
- [Monitoring:Články s identifikátorem GND](#)
- [Monitoring:Články s identifikátorem LCCN](#)
- [Monitoring:Články s identifikátorem LNB](#)
- [Monitoring:Články s identifikátorem NDL](#)
- [Monitoring:Články s identifikátorem NLI](#)

Hledání

Speciální:Hledání

Hledat

Strojový překlad

85 jazyků

[Přidat téma](#)