



TISKOVÁ ZPRÁVA

Univerzita Karlova, Matematicko-fyzikální fakulta
Ke Karlovu 3, 121 16 Praha 2
<https://www.mff.cuni.cz>

Matematicko-fyzikální fakulta vyvinula systém pro automatický překlad mezi češtinou a ukrajinštinou

PRAHA, 16. března 2022 – Výzkumníci z Ústavu formální a aplikované lingvistiky (<https://ufal.mff.cuni.cz>) vyvinuli v rekordním čase veřejně dostupný automatický překladač mezi češtinou a ukrajinštinou, aby pomohli překonat jazykovou bariéru uprchlíků z Ukrajiny v českém prostředí. Překladač provozuje výzkumná infrastruktura LINDAT/CLARIAH-CZ na adrese <https://lindat.cz/translation>.

O uplynulém víkendu skupina dvanácti výzkumníků z Ústavu formální a aplikované lingvistiky MFF UK uspořádala hackathon, jehož cílem bylo zpřístupnit automatický překlad mezi ukrajinštinou a češtinou široké veřejnosti, zejména pro účely informovanosti uprchlíků z Ukrajiny při jejich kontaktu se státní správou, ubytovateli, školami a potenciálními zaměstnavateli. Výsledkem spolupráce s desítkou externích účastníků hackatonu je zjednodušená webová stránka věnovaná jen překladu z a do ukrajinštiny. Předběžné výsledky testů ukazují, že kvalita ukrajinsko-českého překladu je vyšší než např. u systému Google Translate, a to zejména díky novým metodám strojového učení, které původně vznikly pro česko-anglický překlad. Další výhodou pražského systému je, že na rozdíl od jiných volně dostupných online systémů nepoužívá angličtinu jako mezikrok, ale překládá mezi češtinou a ukrajinštinou přímo.

Ústav formální a aplikované lingvistiky se zabývá strojovým překladem již více než 50 let. Za tu dobu se postupně měnily metody a přístupy, až byl překlad pomocí tzv. neuronových sítí vylepšen tak, že v kvalitě anglicko-českého překladu konkuruje překladu lidskému (viz <https://www.mff.cuni.cz/cs/verejnost/aktuality/prekladac-z-matfyzu-dohani-v-kvalite-bezne-prekladatele>). Tento poslední model překladu se nazývá CUBBITT a je dostupný pro několik dvojic jazyků ve výzkumné infrastruktuře LINDAT/CLARIAH-CZ pro výzkumníky, firmy i veřejnost (<https://lindat.cz/services/translation>). V současné době se pracuje na sběru dat i pro další jazyky a zároveň na nových metodách tak, aby se kvalita překladu i nadále zlepšovala. Zároveň další skupiny v rámci Ústavu formální a aplikované lingvistiky pracují – v mezinárodních konsorciích – na modifikacích systému překladu tak, aby umožňoval v budoucnu i tlumočení, tj. překlad mluvené řeči (projekt ELITR, <https://elitr.eu> nebo projekt WELCOME, <https://welcome-h2020.eu>).

Kontakt pro média:

Rudolf Rosa, Martin Popel, Michal Novák, Jindřich Libovický, Jindřich Helcl

(u4u@ufal.mff.cuni.cz, tel.: +420 951 552 954)

Ústav formální a aplikované lingvistiky Matematicko-fyzikální fakulty Univerzity

Karlovy byl založen v roce 1990 jako pokračování výzkumné a pedagogické činnosti bývalé Laboratoře algebraické lingvistiky, existující od počátku 60. let na Filozofické fakultě a později na Matematicko-fyzikální fakultě Univerzity Karlovy. Ústav je především výzkumnou institucí, která se zabývá mnoha tématy v oblasti počítačové lingvistiky a zpracování přirozeného jazyka a která participuje na mnoha výzkumných projektech na národní i mezinárodní úrovni. Je také koordinačním pracovištěm velké výzkumné infrastruktury [LINDAT/CLARIAH-CZ](#), která podporuje výzkum v České republice i ve světě poskytováním jazykových zdrojů, nástrojů a služeb v oblasti jazykových technologií a digitálních humanitních věd. Ústav formální a aplikované lingvistiky nabízí komplexní výukový program jak pro bakalářský a magisterský stupeň (Bc., Mgr.), tak pro doktorské studium (Ph.D.) v oboru počítačové lingvistiky. Všechny programy se vyučují v češtině a angličtině. Institut je také členem double degree magisterského programu LCT EU.

[LINDAT/CLARIAH-CZ](#) (<https://lindat.cz>) je společný distribuovaný český národní uzel celoevropských sítí CLARIN ERIC (Common Language Resources and Technology Infrastructure) a DARIAH ERIC (Digital Research Infrastructure for the Arts and Humanities). Sdružuje 11 výzkumných organizací, které působí v oblasti humanitních věd a umění v České republice – v oblasti lingvistiky, historie a historické bibliografie, kultury a vědy v oblasti kultury, dějin umění, filozofie, filmové kultury, vizuálního umění, muzikologie a dějin hudby, etnologie, folkloru, archeologie a také v některých mezioborových oblastech.