
Proč počítače neumějí česky? Chybí větší databáze

Proč počítače neumějí česky? Chybí větší databáze



Hospodářské noviny, 27.9.2011, Rubrika: Titulní strana, Strana: 1, Autor: Zuzana Keményová

Univerzita Karlova vydala Bílou knihu češtiny. Mapuje problémy s jejím využíváním v informačních technologiích.

Češi používají počítače už přes čtyřicet let, přesto počítače a další přístroje rozumějí česky stále méně. Nedokážou analyzovat text ani porozumět složitějším hlasovým příkazům. Proč? Na to odpovídá nová Bílá kniha češtiny, kterou nyní vydal Ústav formální a aplikované lingvistiky při Matematickofyzikální fakultě Univerzity Karlovy.

„V Česku sice existuje výzkum počítačové lingvistiky, ale stále chybí kompletní jazyková data, se kterými by mohly počítače pracovat a pomocí kterých by se dokázaly logiku české řeči naučit,“ vysvětluje Jan Hajič, jeden z autorů Bílé knihy češtiny, který na fakultě učí počítačovou lingvistiku.

Počítače zatím v češtině zvládají jen základní jazykové technologie. Například opravit překlepy v textovém editoru. „Složitější funkce, jako je kvalitní překlad, porozumění řeči nebo schopnost z česky psaného článku vybrat a vyhodnotit data, to je zatím počítačům cizí,“ dodává Hajič.

Bílá kniha upozorňuje také na fakt, že čeština stejně jako některé další jazyky menších národů mizí z internetu. Lidé se stále více zaměřují na texty psané v angličtině nebo v jiném světovém jazyce. „Když například český lékař hledá něco na internetu, zadá si dotaz raději v angličtině. Vyhledávač jí lépe rozumí, výsledků je víc a jsou přesnější,“ konstatuje Hajič. Pozoruhodné také je, že čeští experti vypracovali studii o tom, že češtinu stále více vytlačují globální jazyky jako angličtina, anglicky. Na českém překladu se teprve pracuje.