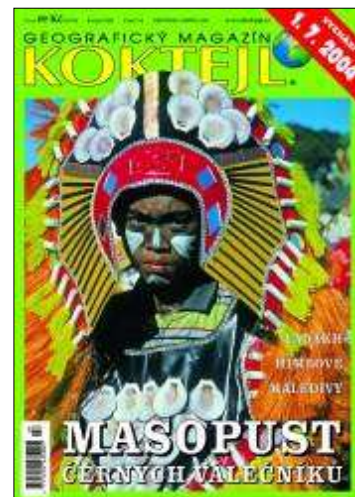




Severočeská vědecká knihovna
v Ústí nad Labem

vás zvou
na jazykový seminář
šéfredaktorky



Dagmar CESTROVÉ

Jak přepsat hokejový zápas

Počítačové lingvisté: vytváříme systém, který by komentář ke sportovnímu utkání automaticky převl do titulků

Může stroj rozumět lidem a plyně hovořit jejich jazykem? O perspektivách a úskalích počítačové lingvistiky s LN hovořil profesor **Josef Psutka** ze Západočeské univerzity v Plzni a docent **Jan Hajič** z Univerzity Karlovy v Praze.

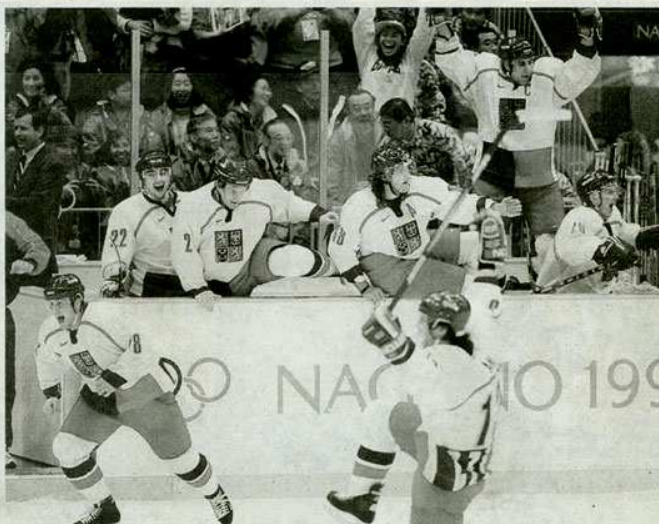
LN: Mohou počítače mluvit lidským hlasem?

Prof. Psutka: Jistě. Dlouho se odborníci snažili o to, aby počítač hovořil srozumitelně. Mluvíme teď o syntéze řeči z psaného textu. Z diskety nahrajete něco do počítače, třeba nějaký novinový článek, a on vám jej přečte. Včetně číslovky, zkratk, cizích jmen a podobně. Problém byl, že posluchači často nerozuměli tomu, co hlasový syntetizér vlastně řekl.

Tyto potíže se ale už podařilo vyřešit. Dnes chceme, aby počítač hovořil co nejpodobněji, aby lidskou řeč napodoboval věrně.

LN: Počítač dokáže nejen hovořit, umí také zachytit, co mu člověk řekne. Proč není ovládání různých přístrojů hlasem více rozšířeno?

Doc. Hajič: Jeden z důvodů je, že počítač vám neporozumí vždy. Splete se třeba jen v jednom ze dvaceti případů. Představte si ale, že byste ovládal hlasem pračku a ona



Počítač převádějící komentář k hokejovému utkání do titulků by musel čelit nejrůznějším situacím - jako byla ta po závěrečné šířené finálového utkání v Naganu, kdy hlas komentátora poněkud nabyl na síle

Foto: CTK

uložena čísla různých lidí. Není problém poručit mu, aby volal třeba Marušku. Kdyby vám ale v jednom z dvaceti případů nerozuměl a vytvořil číslo někoho jiného, asi by se

měl fidič na volantu tlačítka, a kdyby chtěl, aby mu počítač „naslouchal“, stiskl by jej.

LN: Auta určitě nejsou jedinou

dva komentátoři a je možné systém „natrénovat“ přímo na jejich hlas, fráze, slovník. Jde o spontánní projev, a je tedy vyloučeno, aby se vše, co zazní, pře-

LN: Zmínili jste diktování emailů. Jaké je procento chyb?

Doc. Hajič: Uživatele si může přislušný program sám nainstalovat do svého počítače, a pokud má trpělivost, dosáhne slušných výsledků. První týden musí vždy, když se počítač zmýlí, chybu opravit a vložit správnou verzi do paměti. Systém se tak učí rozumět přímo konkrétnímu člověku. Pak dělá jen asi tři až pět procent chyb. Tři až pět slov ze sta napíše špatně.

LN: Existuje něco takového i v češtině?

Prof. Psutka: Myslím, že v současné době ne. Pokud je chyb více než pět procent, zdržují opravy uživatele natolik, že ztrácí zájem.

Slovník Američana nebo Angličana, který diktuje dopis, čítá řádově 20 až 30 tisíc slov. Je v tom obsaženo v podstatě všechno, chybí jen jedno slovo z jednoho či dvou set. Ale v češtině je to jiné. Když budete mít v paměti počítače 60 tisíc slov, pořád bude každé dvacáté nebo pětadvacáté slovo chybět. A říká se, že každé slovo, které není ve slovníku, působí minimálně dvě chyby. Je to principiální problém...

Doc. Hajič: ...který vzniká tím, že čeština je jazyk, kde se slova skládají a časují. Když v této souvislosti říkáme „slovo“, máme na mysli slovní tvar.

LN: Třeba „červení“ a „červení“ jsou dvě různá slova?

Doc. Hajič: Ano. Ale i když mlu-

jete koupit rohlík v Ghaně, to už je mnohem realističtější. Nebo ne v Ghaně, ale třeba v Itálii. Přístroj ve vaší dlaně totiž pak vystačí s omezeným množstvím slov a frází.

Budoucnost určitě patří také automatickému překladu psaného textu. Firmám, které vyvážejí různé zboží, tato pomůcka už dnes šetří peníze. Zvlášť pokud musejí se svými produkty dodávat velké množství dokumentace. Automatický překlad proto hojně využívají třeba výrobci letadel nebo počítačových programů.

LN: Kontroluje přeložený text člověk?

Doc. Hajič: Ano, ale pořád to vyjde levněji než platit armádu překladatelů. Ti navíc nejsou k dispozici, pracovat s technickou dokumentací chce málokdo. Studenti si většinou představují, že budou překládat Shakespeara, ale popíravka je po úplně jinak zaměřených lidech.

Automatický překlad funguje také na internetu. Vyhledávací vám najde článek třeba v čínštině a vy vůbec nemáte potuchy, o čem je řeč. Můžete si jej nechat přeložit — výsledek sice nebude dokonalý, ale aspoň pochopíte, jestli je to o fotbalu, nebo o politice.

Martin Uhlíř

Doc. RNDr. Jan Hajič, Dr., je ředitel ústavu formální a aplikované lingvistiky MFF UK v Praze. Prof. ling. Josef Psutka, CSc., je vedoucí

TAJEMSTVÍ NOVINÁŘSKÉ ČEŠTINY

Jazykový seminář se koná v sále půdní vestavby
Severočeské vědecké knihovny v Ústí nad Labem, W. Churchilla 3
v pondělí 14. dubna 2008 od 8.30 hodin
a ve středu 23. dubna 2008 od 8.30 hodin
vstup volný