

OHROZENIE NAŠICH JAZYKOV: VÝZVA PRE JAZYKOVÉ TECHNOLOGIE

V poslednej dekáde sme svedkami digitálnej revolúcie, ktorá má značný vplyv na komunikáciu a spoločnosť. Nedávne pokroky v digitálnych a sieťových komunikačných technológiách sa niekedy prirovnávajú ku Gutenbergovmu vynájdeniu kníhtlače. Ako nám môže táto analógia konkrétne priblížiť budúcnosť európskej informačnej spoločnosti a našich jazykov?

Sme svedkami digitálnej revolúcie, ktorú môžeme prirovnať ku Gutenbergovmu vynálezu kníhtlače

Po Gutenbergovom vynáleze nastal skutočný prelom v komunikácii a výmene poznatkov vďaka takým snahám, ako bol napr. Lutherov preklad Biblie do zrozumiteľného jazyka. V ďalších storočiach nastal rozvoj kultúrnych postupov, ktoré rozšírili výmenu poznatkov a zefektívnila spracovávanie jazyka. Zmeny, ktoré nastali:

- ortografické a gramatické ustálenie významnejších jazykov umožnilo rýchle rozšírenie nových vedeckých a intelektuálnych ideí;
- rozvoj oficiálnych jazykov pomohol obyvateľom komunikovať v rámci určitých (často politických) hraníc;
- vyučovanie a preklad jazykov umožnil výmenu poznatkov medzi jazykmi;
- vytvorenie žurnalistických a bibliografických príručiek prinieslo zlepšenie kvality a dostupnosti tlačového materiálu;

- vytvorenie rôznych médií, akými sú knihy, noviny, rozhlas, televízia a i. uspokojilo rozmanité komunikačné potreby.

Za posledných dvadsať rokov pomohli informačné technológie automatizovať a uľahčiť celý rad procesov:

- DTP softvér nahradil strojopis a sadzbu;
- prezentačný softvér, ako napríklad Open/Libre-Office Impress alebo Microsoft PowerPoint nahradili spätný projektor;
- zasielanie a prijímanie dokumentov e-mailom je rýchlejšie ako prostredníctvom faxu;
- SIP telefónia alebo Skype umožňujú internetové volania a virtuálne stretnutia;
- efektívne kódovanie zvukových a obrazových súborov uľahčuje výmenu multimediálneho obsahu;
- nástroje na vyhľadávanie umožňujú na báze kľúčových slov efektívny prístup na webové stránky;
- on-line služby, ako napríklad Google Translate, ponúkajú rýchle, aj keď približné preklady;
- platformy sociálnych médií (Pokec, Facebook, Twitter, Google a i.) uľahčujú spoluprácu a sprístupnenie informácií.

Spomenuté nástroje a aplikácie ľuďom pomáhajú, no v súčasnosti nedokážu dostatočne pokryť potreby multilingválnej modernej európskej informačnej spoločnosti, v ktorej je neustály tok informácií a tovaru.

2.1 JAZYKOVÉ HRANICE SPOMAĽUJÚ EURÓPSKU INFORMAČNÚ SPOLOČNOSŤ

V súčasnosti nemôžeme presne odhadnúť, aká bude informačná spoločnosť o niekoľko rokov. Je však veľmi pravdepodobné, že revolúcia v komunikačných technológiách spojí ľudí, ktorí hovoria rozličnými jazykmi, napriek jazykovým bariéram. Momentálne môžeme cítiť istý tlak na ľudí, aby sa učili cudzie jazyky, a najmä na ľudí, ktorí by mali vytvárať nové technologické aplikácie na zabezpečenie vzájomného dorozumenia. V aktuálnej globálnej ekonomike a informačnom priestore sa denne konfrontujeme s narastajúcim počtom jazykov, hovoriacimi a novými témami. Súčasná popularita sociálnych médií (Wikipedia, Facebook, Twitter, YouTube, Pôkec, Google+) je len špičkou tohto pokrokového ľadovca.

V globálnej ekonomike a informačnom priestore sa denne konfrontujeme s rôznymi jazykmi, hovoriacimi a novými témami

Dnes dokážeme prenášať gigabajty textu po celom svete za pár sekúnd, hoci sú v jazyku, ktorému nerozumieme. Podľa nedávnej správy, ktorú vydala Európska komisia, 57 % používateľov internetu platí za tovar a služby v cudzom jazyku (angličtina je najbežnejšia, hneď za ňou nasleduje francúzština, nemčina a španielčina). 55 % používateľov číta obsah v cudzom jazyku, pričom iba 35 % používa iný jazyk na písanie e-mailov alebo posielanie komentárov na webe [2]. Pred niekoľkými rokmi mohla byť angličtina internetová lingua franca, pretože prevažná väčšina materiálov na webe bola v angličtine. Situácia sa však medzičasom modifikovala – rozrástlo sa množstvo inojazyčného on-line obsahu (najmä ázijského a arabského).

Táto digitálna priepať, ktorá je zapríčinená jazykovými bariérami, prekvapivo nezískala dostatok pozornosti na

verejnosti. Digitálny svet si kladie naliehavú otázku: „Ktorým európskym jazykom sa bude dariť v zosieťovanej informačnej a znalostnej spoločnosti a ktoré zaniknú?“

2.2 NAŠE JAZYKY V OHROZENÍ

Kníhtlať značne prispela k výmene informácií v Európe, ale napomohla tiež zániku mnohých európskych jazykov. V regionálnych a menšinových jazykoch sa dokumenty rozmnožovali zriedkakedy. Výsledkom bolo, že mnohé jazyky, ako napríklad rómsky alebo rusínsky, sa zredukovali viacmenej len na ústne podanie, čo obmedzovalo ich kontinuálne osvojenie a rozšírenie. Bude mať internet podobný vplyv aj na naše jazyky?

Rôznorodosť jazykov v Európe je súčasťou kultúrneho bohatstva Európy

Približne 80 jazykov je časťou najvzácnejšieho a najdôležitejšieho kultúrneho bohatstva Európy. Množstvo európskych jazykov je takisto nevyhnutnou súčasťou jej sociálneho úspechu [3]. Zatiaľ čo sa budú populárne jazyky ako angličtina a španielčina v rozvíjajúcej sa digitálnej spoločnosti a na trhu určite udržiavať, mnohé európske jazyky sa vynechajú z digitálnych komunikácií a pre internetovú spoločnosť sa stanú irelevantné. Takýto vývoj by oslabil európsku stabilitu, pretože by bol v rozpore s cieľom zabezpečiť rovnaké postavenie každého európskeho občana bez ohľadu na jazykovú príslušnosť. V správe Unesca o multilingvizme sa uvádza, že jazyky sú médium uplatňovania základných ľudských práv, ako je právo na vyjadrenie politického názoru, vzdelanie a účasť na spoločenskom živote [4].

2.3 JAZYKOVÉ TECHNOLOGIE SÚ KLÍČOVÝMI TECHNOLOGIAMI

V minulosti sa najviac investovalo do jazykového vzdelávania a prekladu. Podľa niektorých odhadov sa napríklad v roku 2008 v Európe minulo na preklad, interpretáciu, softvérovú lokalizáciu a internetovú globalizáciu približne 8,4 miliardy eur, pričom sa rátať s 10-percentným nárastom ročne [5]. Faktom je, že tieto finančné prostriedky napriek tomu nestačia na uspokojenie súčasných ani budúcich potrieb. Najlepšie riešenie pre dostatočný výskum používania jazyka je výber technológií, ktorú používame aj na riešenie problémov v doprave, energetike, sociálnej oblasti a pod.

Digitálne jazykové technológie (v písanom aj hovorenom diskurze) pomáhajú ľuďom spolupracovať, podnikáť, sprístupňovať vedomosti a zúčastňovať sa na sociálnych a politických diskusiách bez ohľadu na jazykové bariéry alebo počítačové zručnosti. Sú užitočné v prípade:

- vyhľadávania informácií pomocou internetového vyhľadávača,
- kontroly pravopisu a gramatiky v textových procesoroch,
- odporúčania produktu v internetovom obchode,
- počúvania inštrukcií automobilového navigačného systému,
- prekladu webových stránok prostredníctvom online služieb.

Jazykové technológie sa skladajú z niekoľkých základných aplikácií, ktoré sú bázou väčšieho aplikačného rámca. Účelom bielej knihy META-NET-u je preskúmať stav základných technológií všetkých európskych jazykov.

Aby si Európa udržala svoju pozíciu na čele inovatívneho pokroku, mali by sa jazykové technológie adap-

tovať dôkladne a cenovo dostupne na všetky európske jazyky a zároveň sa pevne integrovať do kľúčových softvérových prostredí. Bez jazykových technológií Európa nedosiahne efektívne, interaktívne, multimedálne a viacjazyčné používateľské prostredie.

Európa potrebuje vhodné a cenovo dostupné jazykové technológie pre všetky európske jazyky

2.4 PRÍLEŽITOSTI PRE JAZYKOVÉ TECHNOLOGIE

V oblasti tlače bolo technologickým zlomom vynájdenie tlačiarne. Ľudia sa namáhali pri prácnom vyhľadávaní, čítaní, prekladaní a sumarizácii poznatkov. Čakali sme až na Edisona, ktorý zachytil hovorenú reč, a jeho technológia vytvárala stále iba analógové kópie.

Digitálne jazykové technológie dokážu vytvoriť automatický preklad, vygenerovať obsah, spracúvať informácie a riadiť vedomostný manažment, ktorý je aplikovateľný na všetky európske jazyky. Jazykové technológie môžu tiež podporovať rozvoj používateľských rozhraní pre domácu elektroniku, zariadenia, dopravné prostriedky, počítače či roboty. Hoci existuje mnoho takýchto prototypov, komerčné a priemyselné aplikácie sú stále iba v prvotných štádiách rozvoja. Nedávne úspechy vo výskume a rozvoji vytvorili skutočný priestor na nové možnosti. Povedzme strojový preklad je už primerane presný v špecifických oblastiach; experimentálne aplikácie poskytujú mnohojazyčnú informáciu a vedomostný manažment, ako aj generovanie obsahu v mnohých európskych jazykoch.

Ako pri väčšine technológií, aj prvé jazykové aplikácie, ako napríklad hlasové používateľské rozhrania a dialógové systémy, boli vyvinuté pre vysoko špecializované domény a často vykazujú obmedzenú použiteľnosť. Ale v oblasti vzdelávania a zábavného priemyslu sú obrovské príležitosti na integráciu jazykových technológií do

hier, edukačných pomôcok, simulačných prostredí, prípadne vzdelávacích programov. Mobilné informačné služby, softvéry na počítačovú podporu učenia sa jazyka, e-learningové prostredia, nástroje na sebahodnotenie a softvéry na detekciu plagiátorstva sú len zlomkom možností, v ktorých zohrávajú jazykové technológie dôležitú úlohu. Popularita sociálnych aplikácií ako Twitter, Pokec alebo Facebook naznačuje potrebu sofistikovanejších jazykových technológií, ktoré dokážu monitorovať príspevky, sumarizovať diskusie, navrhnúť názorové trendy, detegovať emocionálne reakcie, identifikovať porušenie autorských práv alebo vystopovať zneužitie diela.

Jazykové technológie môžu pomôcť prekonať bariéry lingvistickej rozmanitosti

Jazykové technológie predstavujú pre Európsku úniu obrovskú príležitosť. Môžu pomôcť pri problematike viacjazyčnosti v Európe – keďže obchodná sféra, rôzne organizácie či školy sú charakteristické svojou národnostnou rozmanitosťou. Jazykové technológie môžu pomôcť prekonať jazykové bariéry vďaka slobodnému a otvorenému používaniu rozličných jazykov. Pri pohľade na budúcnosť nám zavedenie inovatívnych a multilingválnych jazykových technológií pre Európu takisto môže pomôcť v komunikácii s celosvetovými partnermi a s ich viacjazyčnými spoločenstvami. Jazykové technológie možno vnímať aj ako „podporné“ prostriedky, ktoré prekonávajú jazykovú rozmanitosť a zbližujú jazykové spoločenstvá.

Napokon, jedno odvetvie výskumu predstavuje aj používanie jazykových technológií pri záchranných akciách v oblastiach postihnutých katastrofami, kde ich použitie môže byť otázkou života a smrti, napríklad budúce inteligentné roboty s mnohorakými jazykovými schopnosťami majú potenciál zachraňovať ľudské životy.

2.5 VÝZVY PRE JAZYKOVÉ TECHNOLOGIE

Hoci jazykové technológie za posledné roky napredujú, súčasné tempo technologického vývoja a inovácie produktov je pomalé. Jazykové technológie so širokým využitím (napríklad kontrola pravopisu a gramatiky v textových editoroch) jestvujú v monolingválnej forme, a preto sú dostupné len pre hŕstku jazykov. On-line služby, ako sú profesionálne aplikácie strojových prekladov, prinášajú so sebou mnohé ťažkosti v situáciách, v ktorých sú potrebné veľmi presné a úplné preklady. Vzhľadom na zložitosť ľudského jazyka a modelovanie nášho jazyka do softvéru je následné testovanie prídlhé a nákladné a vyžaduje si neustálu finančnú podporu. Ak si chce Európa zachovať svoje postavenie priekopníka v prijímaní technologických výziev viacjazyčnej jazykovej komunity, musí neustále predkladať nové metódy na urýchlenie technologického rozvoja, napríklad progres v oblasti počítačovej technológie a techník ako crowd-sourcing.

Súčasné tempo technologického vývoja je príliš pomalé

2.6 OSVOJOVANIE SI JAZYKA

Aby sme si vedeli lepšie predstaviť prácu počítača s osvojením si jazyka, stručne zhrnieme spôsoby, akými si ľudia osvojujú prvý a druhý jazyk. Potom si načrtujeme, ako si jazyk osvojujú jazykové technológie.

Ľudia si jazyk osvojujú dvoma rozličnými spôsobmi. V prvom prípade sa dieťa učí jazyk tak, že počúva rozhovory medzi hovoriacimi v danom jazyku. Presnejšie, jazykovými vzormi sú preňho používatelia jazyka, ako napríklad rodičia, súrodenci alebo iní rodinní príslušníci. Dieťa začína produkovať prvé slová a krátke frázy

vo veku približne dvoch rokov. Deje sa to vďaka špeciálnej genetickej dispozícii imitovať zvuky a následne si odôvodniť to, čo počuje.

Učenie sa druhého jazyka zvyčajne vyžaduje oveľa viac úsilia, lebo dieťa už nie je súčasťou jazykového spoločenstva rodených hovoriacich. V školskom veku sa cudzie jazyky väčšinou osvojujú učením gramatických štruktúr, slovnej zásoby a pravopisu z kníh a vzdelávacích materiálov, ktoré opisujú jazykové systémy pomocou abstraktných pravidiel, tabuliek a textových ukážok. Učenie sa cudzieho jazyka si vyžaduje veľa času i úsilia a s pribúdajúcim vekom to už nie je také jednoduché.

Jazykové technológie nadobúdajú jazykové schopnosti podobným spôsobom ako ľudia. Štatistické prístupy získavajú jazykové schopnosti z rozmanitého výberu konkrétnych príkladov textov. Tieto algoritmy strojového učenia modelujú istý druh jazykovej schopnosti, ktorá dokáže odvodzovať vzory ako slová, krátke frázy a celé vety používané v jednom jazyku alebo prekladané z jedného jazyka do druhého.

Ľudia si osvojujú jazyk pozorovaním komunikácie a učením sa jazykových pravidiel

Tento štatistický prístup vyžaduje obsah miliónov viet a svoj kvalitatívny výkon zvyšuje s narastajúcim množstvom analyzovaných textov. To je jeden z dôvodov, prečo sa prevádzkovatelia vyhľadávačov snažia získať čo najviac písomných materiálov. Korekcia pravopisu v textových procesoroch a služby ako Google Hľadať na webu (oficiálny názov služby) a Google Translate sú závislé od štatistických prístupov. Veľkou výhodou štatis-

tiky je, že stroj sa učí veľmi rýchlo, hoci kvantita nie vždy korešponduje s kvalitou.

Systémy založené na pravidlách sú druhým najväčším typom jazykových technológií. Vysoko špecializovaní odborníci z oblasti lingvistiky, počítačovej lingvistiky a počítačovej vedy kódujú gramatické analýzy (pravidlá prekladu) a zostavujú zoznam slovnej zásoby (lexikóny). Vytvorenie týchto systémov je časovo náročné a prácne. Niektoré z týchto hlavných systémov strojového prekladu založených na pravidlách sa rozvíjajú už viac než 20 rokov. Ich výhodou je, že odborní pracovníci môžu systematickejšie kontrolovať spracúvanie jazyka, čo prispieva k oprave prípadných chýb v softvéri. Vďaka týmto systémom sa používateľovi poskytne detailnejšia spätná väzba, osobitne vtedy, keď sa tieto systémy používajú na výučbu jazykov. Z finančných dôvodov sú systémy založené na pravidlách prístupné iba pre rozšírenejšie jazyky. Silné a slabé stránky štatistických systémov a systémov založených na pravidlách sa navzájom dopĺňajú. Aktuálny výskum sa sústreďuje na hybridné prístupy, ktoré tieto dva systémy kombinujú. Doteraz sa však viac uplatnili v priemyselných aplikáciách než v oblasti výskumu. Ako sme si v tejto kapitole mohli prečítať, v dnešnej informačnej spoločnosti sa využíva množstvo jazykových technológií. Kvôli viacjazyčnosti to platí najmä pre európsky ekonomický a informačný priestor. Jazykové technológie zaznamenali v posledných rokoch značný rozmach. Ich permanentné zdokonaľovanie však je nevyhnutnosťou.

V nasledujúcich kapitolách opíšeme úlohu slovenského jazyka v európskej informačnej spoločnosti a zhodnotíme súčasný stav jazykových technológií pre slovenský jazyk.

The Slovak Language in the Digital Age

Rehm, G.; Uszkoreit, H. (Eds.)

2012, VI, 85 p. 25 illus., 23 illus. in color., Softcover

ISBN: 978-3-642-30369-2