META NET

White Paper Series Séria bielych kníh

THE SLOVAK LANGUAGE IN THE DIGITAL AGE

SLOVENSKÝ JAZYK V DIGITÁLNOM VEKU

Mária Šimková
Radovan Garabík
Katarína Gajdošová
Michal Laclavík
Slavomír Ondrejovič
Jozef Juhár
Ján Genči
Karol Furdík
Helena Ivoríková
Jozef Ivanecký

Springer

White Paper Series Séria bielych kníh

THE SLOVAK LANGUAGE IN THE DIGITAL AGE

SLOVENSKÝ JAZYK V DIGITÁLNOM VEKU

Mária Šimková Jazykovedný ústav Ľ. Štúra SAV
Radovan Garabík Jazykovedný ústav Ľ. Štúra SAV
Katarína Gajdošová Jazykovedný ústav Ľ. Štúra SAV
Michal Laclavík Ústav informatiky SAV
Slavomír Ondrejovič Jazykovedný ústav Ľ. Štúra SAV
Jozef Juhár Technická univerzita v Košiciach
Ján Genči Technická univerzita v Košiciach
Karol Furdík Technická univerzita v Košiciach
Helena Ivoríková Studia Academica Slovaca UK
Jozef Ivanecký European Media Laboratory

Georg Rehm, Hans Uszkoreit
(redakcia, editors)

Editors
Georg Rehm
DFKI
Alt-Moabit 91c
Berlin 10559
Germany
e-mail: georg.rehm@dfki.de

Hans Uszkoreit
DFKI
Alt-Moabit 91c
Berlin 10559
Germany
e-mail: hans.uszkoreit@dfki.de

ISSN 2194-1416 ISSN 2194-1424 (electronic)
ISBN 978-3-642-30369-2 ISBN 978-3-642-30370-8 (eBook)
DOI 10.1007/978-3-642-30370-8
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012940340

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

PREDHOVOR PREFACE

Táto biela kniha je súčasťou série, ktorá propaguje najnovšie poznatky a potenciál jazykových technológií. Je určená novinárom, politikom, jazykovým spoločnostiam, učiteľom a iným. V európskych krajinách majú jazykové technológie rozličnú úroveň aj využitie. Z toho dôvodu sú aj opatrenia potrebné na ďalšiu podporu výskumu a vývoja jazykových technológií pre každý jazyk odlišné. Požadované opatrenia závisia od mnohých faktorov, akými sú napríklad zložitosť daného jazyka či veľkosť jazykovej komunity.

META-NET, sieť excelentnosti, financovaná z fondov Európskej komisie, vypracovala v tejto sérii bielych kníh (s. 85) analýzu súčasných jazykových zdrojov a technológií. Analýza zahŕňala okrem 23 oficiálnych európskych jazykov aj iné dôležité národné i regionálne jazyky Európy. Výsledky analýzy poukázali na značné nedostatky v technologickej podpore a na medzery vo výskume pre každý jazyk. Podrobnejšia expertná analýza a zhodnotenie momentálnej situácie pomôže maximalizovať efektivitu ďalších výskumov.

Od novembra 2011 META-NET pozostáva z 54 výskumných centier v 33 krajinách Európy (s. 81). META-NET spolupracuje so zainteresovanými stranami z oblasti ekonómie (softvérové spoločnosti, poskytovatelia technológií a používatelia), z oblasti vládnych agentúr, výskumných organizácií, nevládnych organizácií, jazykových spoločenstiev a európskych univerzít. META-NET spoločne s týmito komunitami vytvára jednotnú technologickú víziu a strategický plán výskumu pre multilingválnu Európu 2020.

This white paper is part of a series that promotes knowledge about language technology and its potential. It addresses journalists, politicians, language communities, educators and others. The availability and use of language technology in Europe varies between languages. Consequently, the actions that are required to further support research and development of language technologies also differ. The required actions depend on many factors, such as the complexity of a given language and the size of its community.

META-NET, a Network of Excellence funded by the European Commission, has conducted an analysis of current language resources and technologies in this white paper series (p. 85). The analysis focuses on the 23 official European languages as well as other important national and regional languages in Europe. The results of this analysis suggest that there are tremendous deficits in technology support and significant research gaps for each language. The given detailed expert analysis and assessment of the current situation will help maximise the impact of future research.

As of November 2011, META-NET consists of 54 research centres in 33 European countries (p. 81). META-NET is working with stakeholders from economy (software companies, technology providers and users), government agencies, research organisations, non-governmental organisations, language communities and European universities. Together with these communities, META-NET is creating a common technology vision and strategic research agenda for multilingual Europe 2020.

Autori tohto dokumentu ďakujú autorom Bielej knihy pre nemčinu za povolenie používať vybrané jazykovo nezávislé materiály z ich dokumentu [1].

Táto biela kniha bola financovaná prostredníctvom Siedmeho rámcového programu a Programu podpory politiky v oblasti informačných a komunikačných technológií Európskej komisie na základe dohôd T4ME (Grantová dohoda 249119), CESAR (Grantová dohoda 271022), METANET4U (Grantová dohoda 270893) a META-NORD (Grantová dohoda 270899).

The authors of this document are grateful to the authors of the White Paper on German for permission to re-use selected language-independent materials from their document [1].

The development of this White Paper has been funded by the Seventh Framework Programme and the ICT Policy Support Programme of the European Commission under the contracts T4ME (Grant Agreement 249119), CESAR (Grant Agreement 271022), METANET4U (Grant Agreement 270893) and META-NORD (Grant Agreement 270899).

OBSAH CONTENTS

SLOVENSKÝ JAZYK V DIGITÁLNOM VEKU

THE SLOVAK LANGUAGE IN THE DIGITAL AGE

ZHRNUTIE

Európa sa počas posledných 60 rokov stala významnou politickou a ekonomickou silou, kultúrne a jazykovo je však stále veľmi rôznorodá. To znamená, že od Portugalska po Poľsko a od Talianska po Island je bežná komunikácia medzi občanmi Európy podobne ako komunikácia v oblasti podnikania a politiky neustále komplikovaná kvôli jazykovým bariéram. Európske inštitúcie minú ročne približne miliardu eur na preklady inojazyčných textov a na tlmočenie. Nemuselo by to tak byť, ak by moderné jazykové technológie a lingvistický výskum pomohli prekonať jazykové hranice. Ak vhodne využijeme inteligentné zariadenia a aplikácie, budeme môcť navzájom diskutovať alebo obchodovať a rôznosť jazykov nebude pre nás prekážkou.

Jazykové technológie predstavujú mosty

Jedným zo spôsobov, ako prekonať jazykové bariéry, je naučiť sa niekoľko cudzích jazykov. Zvládnuť 23 oficiálnych jazykov členských štátov EÚ a približne 60 ďalších európskych jazykov je však málo pravdepodobné. Vďaka technologickej podpore už dokážeme viesť politické aj ekonomické rokovania, ako aj napredovať vo výskume. Riešením mnohojazyčnosti je vybudovanie kľúčových technológií, ktoré európskym činiteľom ponúknu obrovské výhody, a to nielen v rámci spoločného európskeho trhu, ale aj pri obchodných vzťahoch s krajinami tretieho sveta, najmä s krajinami rozvíjajúcich sa ekonomík. Aby sme dosiahli tento cieľ a zároveň zachovali kultúrnu a jazykovú rozmanitosť, musíme systematicky analyzovať špecifiká všetkých európskych jazykov, ako aj stav súčasných jazykových technológií. Navrhnuté riešenia budú mostom medzi jazykmi.

Jazykové technológie sú kľúčom do budúcnosti

Rozvoj jazykových technológií pre slovenčinu a počítačového spracovania slovenského jazyka v porovnaní so susednými krajinami značne zaostáva. Napríklad v Českej republike sa výskum spracovania prirodzeného jazyka realizuje od polovice 90. rokov minulého storočia a zároveň tu majú jazykové technológie silnú komerčnú podporu. Za prvý významný krok rozvoja jazykových technológií sa na Slovensku považuje vybudovanie Slovenského národného korpusu na začiatku 21. storočia. Prvé veľké projekty zamerané na jazykové technológie a zdroje na Slovensku boli osobitne schválené a financované vládou. Išlo o projekty *Vybudovanie Národného korpusu slovenského jazyka a elektronizácia jazykovedného výskumu v rokoch 2002 – 2006* a *Komplexné spracovanie slovenského jazyka a jeho elektronizácia na účely jazykovedného výskumu*. Obidva projekty sa realizovali v Jazykovednom ústave Ľudovíta Štúra Slovenskej akadémie vied. Projekt ďalej pokračoval pod názvom *Budovanie Slovenského národného korpusu a elektronizácia jazykovedného výskumu na Slovensku (druhá etapa)* na základe zmluvy o jeho spolufinancovaní medzi Ministerstvom školstva SR, Ministerstvom kultúry SR a SAV. Ďalším významným projektom v spracovaní slovenského jazyka bol projekt *APD – Automatický prepis diktátu pre Ministerstvo spravodlivosti Slovenskej republiky* koordinovaný Oddelením analýzy a syntézy reči Ústavu

informatiky Slovenskej akadémie vied v spolupráci s Katedrou elektroniky a multimediálnych komunikácií Technickej univerzity v Košiciach, realizovaný v rokoch 2009–2011. Cieľom bolo vytvoriť systém na prepis hovoreného slovenského jazyka, špeciálne v oblasti súdnictva. Projekt bol financovaný Ministerstvom spravodlivosti SR. V súčasnosti sa systém začína využívať v pilotnej prevádzke na súdoch Slovenskej republiky.

Tieto projekty sú na Slovensku doteraz jedinou významnou iniciatívou v oblasti počítačového spracovania slovenčiny. Ako uvádza naša séria bielych kníh, úroveň výskumu a stavu jazykových technológií je na Slovensku v porovnaní s inými európskymi krajinami oveľa nižšia. Preto je nevyhnutné zvýšiť úroveň jazykových technológií pre slovenčinu.

Dlhodobým cieľom META-NET-u je poskytnúť kvalitné jazykové technológie všetkým jazykom, aby sa napriek kultúrnym rozdielom dosiahla politická a ekonomická jednota. Technologické nástroje pomôžu prekonať existujúce bariéry. Všetky zainteresované strany (z oblasti politiky, vedy, obchodu a pod.) by sa mali snažiť o zjednotenie.

Jazykové technológie pomáhajú zjednotiť Európu

Séria bielych kníh dopĺňa aj ďalšie aktivity META-NET-u (pozri prílohu). Aktuálne informácie, napríklad najnovšie vízie alebo strategický výskumný program META-NET-u, sú dostupné na oficiálnej webovej stránke META-NET-u: http://www.meta-net.eu.