
White Paper Series

Séria bielych kníh

THE SLOVAK
LANGUAGE IN
THE DIGITAL
AGE

SLOVENSKÝ
JAZYK
V DIGITÁLNO
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

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

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, nevládných organizácií, jazykových spoločností 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 DIGITÁLNO M VEKU

1	Zhrnutie	1
2	Ohrozenie našich jazykov: Výzva pre jazykové technológie	3
2.1	Jazykové hranice spomaľujú európsku informačnú spoločnosť	4
2.2	Naše jazyky v ohrození	4
2.3	Jazykové technológie sú kľúčovými technológiami	5
2.4	Príležitosti pre jazykové technológie	5
2.5	Výzvy pre jazykové technológie	6
2.6	Osvojovanie si jazyka	6
3	Slovenčina v európskej informačnej spoločnosti	8
3.1	Všeobecné fakty	8
3.2	Špecifiká slovenčiny	11
3.3	Slovenčina na internete	12
3.4	Slovenčina ako cudzí jazyk	13
3.5	Slovenský národný korpus	15
4	Jazykové technológie na podporu slovenčiny	17
4.1	Architektúra aplikácií	17
4.2	Základné aplikačné oblasti	19
4.3	Ďalšie aplikačné oblasti	27
4.4	Jazykové technológie vo vzdelávaní	29
4.5	Štátne programy a iniciatívy	29
4.6	Dostupnosť nástrojov a zdrojov	30
4.7	Porovnanie jazykov	32
4.8	Záver y	33
5	O META-NET-e	37

THE SLOVAK LANGUAGE IN THE DIGITAL AGE

1	Executive Summary	39
2	Languages at Risk: a Challenge for Language Technology	41
2.1	Language Borders Hold back the European Information Society	42
2.2	Our Languages at Risk	42
2.3	Language Technology is a Key Enabling Technology	42
2.4	Opportunities for Language Technology	43
2.5	Challenges Facing Language Technology	44
2.6	Language Acquisition in Humans and Machines	44
3	Slovak in the European Information Society	46
3.1	General Facts	46
3.2	Particularities of the Slovak Language	49
3.3	Slovak on the Internet	51
3.4	Slovak as a Foreign Language	51
3.5	Slovak National Corpus	53
4	Language Technology Support for Slovak	55
4.1	Application Architectures	55
4.2	Core Application Areas	57
4.3	Other Application Areas	65
4.4	Language Technology in Education	66
4.5	National Projects and Initiatives	67
4.6	Availability of Tools and Resources	67
4.7	Cross-language Comparison	70
4.8	Conclusions	70
5	About META-NET	74
A	Zoznam literatúry – References	75
B	Členovia META-NET-u – META-NET Members	81
C	Séria bielych kníh META-NET-u – The META-NET White Paper Series	85