
White Paper Series

Weißbuch-Serie

THE GERMAN LANGUAGE IN THE DIGITAL AGE

DIE DEUTSCHE SPRACHE IM DIGITALEN ZEITALTER

Aljoscha Burchardt DFKI

Markus Egg Humboldt-Universität zu Berlin

Kathrin Eichler DFKI

Brigitte Krenn ÖFAI

Jörn Kreutel FH Brandenburg

Annette Leßmöllmann Hochschule Darmstadt

Georg Rehm DFKI

Manfred Stede Universität Potsdam

Hans Uszkoreit Universität des Saarlandes, DFKI

Martin Volk Universität Zürich

Georg Rehm, Hans Uszkoreit

(Herausgeber, 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 978-3-642-27165-6 ISBN 978-3-642-27166-3 (eBook)
DOI 10.1007/978-3-642-27166-3
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012947391

© 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)



VORWORT

Dieses Weißbuch gehört zu einer Serie, die Wissen über Sprachtechnologie und deren Potenzial vermitteln soll und sich insbesondere an Journalisten, Politiker, Sprachgemeinschaften und Lehrende richtet. Die derzeitige Verfügbarkeit und Nutzung von Sprachtechnologie in Europa variiert stark je nach Sprache. Folglich müssen auch die notwendigen Maßnahmen für die künftige Unterstützung dieses Bereiches durch Forschung und Entwicklung variieren. Die Maßnahmen hängen von zahlreichen Faktoren ab, z. B. der Komplexität einer Sprache und der Anzahl ihrer Sprecher. META-NET, ein von der Europäischen Kommission gefördertes Spitzenforschungsnetzwerk, hat die aktuellen Sprachressourcen und -technologien in der vorliegenden Weißbuch-Serie analysiert (siehe S. 81). Die Analyse umfasst die 23 europäischen Amtssprachen sowie weitere wichtige nationale und regionale Sprachen Europas. Die Ergebnisse zeigen, dass für alle Sprachen beträchtliche Defizite in der technologischen Unterstützung sowie signifikante Forschungslücken existieren. Die ausführliche Expertenanalyse und Bewertung der aktuellen Situation dient dazu, die Wirksamkeit künftiger Forschung zu maximieren.

META-NET besteht aus 54 Forschungszentren in 33 Ländern (siehe S. 77), die mit Interessensvertretern aus Wirtschaft (Softwareunternehmen, Technologieanbietern, Nutzern), Verwaltung, NGOs, Sprachgemeinschaften und europäischen Universitäten zusammenarbeiten. Zusammen mit diesen Gruppen entwickelt META-NET eine gemeinsame Vision für das Technologiegebiet und eine strategische Forschungsagenda für das mehrsprachige Europa 2020.

PREFACE

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. 81). 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 to maximise the impact of future research.

META-NET consists of 54 research centres in 33 European countries [1] (p. 77). 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.

Die Erstellung dieses Weißbuchs wurde mit Mitteln aus dem Siebten Rahmenprogramm und dem Programm zur Unterstützung der Politik für Informations- und Kommunikationstechnologien der Europäischen Kommission im Rahmen der Verträge T4ME (Finanzhilfvereinbarung 249 119), CESAR (Finanzhilfvereinbarung 271 022), METANET4U (Finanzhilfvereinbarung 270 893) und META-NORD (Finanzhilfvereinbarung 270 899) finanziert.

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 249 119), CESAR (Grant Agreement 271 022), METANET4U (Grant Agreement 270 893) and META-NORD (Grant Agreement 270 899).



INHALTSVERZEICHNIS TABLE OF CONTENTS

DIE DEUTSCHE SPRACHE IM DIGITALEN ZEITALTER

1	Zusammenfassung	1
2	Unsere Sprachen in Gefahr: Eine Herausforderung für die Sprachtechnologie	4
2.1	Sprachgrenzen bremsen die Europäische Informationsgesellschaft	5
2.2	Unsere Sprachen in Gefahr	5
2.3	Sprachtechnologie ist eine Schlüsseltechnologie	6
2.4	Chancen für die Sprachtechnologie	6
2.5	Herausforderungen für die Sprachtechnologie	7
2.6	Spracherwerb bei Mensch und Maschine	8
3	Deutsch in der europäischen Informationsgesellschaft	10
3.1	Allgemeine Fakten	10
3.2	Besonderheiten der deutschen Sprache	11
3.3	Jüngste Entwicklungen	12
3.4	Sprachkultivierung in Deutschland	13
3.5	Sprache im Bildungswesen	14
3.6	Internationale Aspekte	15
3.7	Deutsch im Internet	16
4	Sprachtechnologie für das Deutsche	17
4.1	Anwendungsarchitekturen	17
4.2	Zentrale Anwendungsbereiche	19
4.3	Andere Anwendungsbereiche	27
4.4	Studiengänge und Bildungsprogramme	29
4.5	Nationale Projekte und Initiativen	30
4.6	Verfügbarkeit von Tools und Ressourcen	31
4.7	Sprachübergreifender Vergleich	33
4.8	Schlussfolgerungen	34
5	Über META-NET	38

THE GERMAN LANGUAGE IN THE DIGITAL AGE

1	Executive Summary	39
2	Languages at Risk: a Challenge for Language Technology	42
2.1	Language Borders Hold Back the European Information Society	43
2.2	Our Languages at Risk	43
2.3	Language Technology is a Key Enabling Technology	44
2.4	Opportunities for Language Technology	44
2.5	Challenges Facing Language Technology	45
2.6	Language Acquisition in Humans and Machines	45
3	The German Language in the European Information Society	47
3.1	General Facts	47
3.2	Particularities of the German Language	48
3.3	Recent Developments	49
3.4	Official Language Protection in Germany	49
3.5	Language in Education	50
3.6	International Aspects	51
3.7	German on the Internet	52
4	Language Technology Support for German	54
4.1	Application Architectures	54
4.2	Core Application Areas	55
4.3	Other Application Areas	63
4.4	Educational Programmes	64
4.5	National Projects and Initiatives	64
4.6	Availability of Tools and Resources	66
4.7	Cross-Language Comparison	67
4.8	Conclusions	68
5	About META-NET	72
A	Literaturverweise – References	73
B	META-NET Mitglieder – META-NET Members	77
C	Die META-NET Weissbuch-Serie – The META-NET White Paper Series	81