

## Mindestens 21 europäische Sprachen vom digitalen Aussterben bedroht

### Gute und schlechte Nachrichten zum Europäischen Tag der Sprachen

Die meisten europäischen Sprachen werden in der digitalen Welt unter Umständen nicht überleben, warnt eine Studie der führenden europäischen Sprachtechnologie-Experten. Zu diesem Ergebnis gelangt die Studie „Europe’s Languages in the Digital Age“, in der Wissenschaftler den Stand der Sprachtechnologie für 30 der ca. 80 europäischen Sprachen bewertet haben. Sie kommen zu dem Schluss, dass bei 21 der 30 untersuchten Sprachen ein digitaler Rückhalt nicht existiert oder bestenfalls schwach ist. Durchgeführt wurde die Studie von META-NET, einem europäischen Exzellenz-Netzwerk, in dem 60 Spitzenforschungszentren in 34 Ländern an den technologischen Grundlagen für ein multilinguales Europa arbeiten.

Europa muss Maßnahmen ergreifen, um seine Sprachen auf das digitale Zeitalter vorzubereiten. Sie sind ein wertvoller Bestandteil unseres kulturellen Erbes und als solche verdienen sie Zukunftssicherheit. Am 26. September wird auch dieses Jahr der Europäische Tag der Sprachen stattfinden und an dieses sprachliche und kulturelle Erbe erinnern, dessen Pflege und Förderung für Europa einen wichtigen Stellenwert besitzt. Die META-NET-Studie zeigt, dass das multilinguale Europa im digitalen Zeitalter vor vielen Herausforderungen steht, aber auch immense Chancen hat.

Die Studie, erstellt von mehr als 200 Experten aus Wissenschaft und Industrie, dokumentiert und beurteilt in den 30 Bänden der META-NET White Paper Serie (sowohl online als auch in gedruckter Form erschienen) für jede der untersuchten Sprachen die sprachtechnologische Unterstützung in vier Anwendungsgebieten: automatische Übersetzung, Erkennung und Generierung gesprochener Sprache sowie Textanalyse und Verfügbarkeit von Sprachressourcen. Insgesamt 21 der 30 Sprachen (70%) wurden von den Experten mindestens einmal in die schlechteste Kategorie „Unterstützung ist schwach oder nicht-existent“ eingeordnet. Mehrere Sprachen, zum Beispiel Isländisch, Litauisch und Maltesisch, erhalten diese Bewertung in allen vier Gebieten. Nur das Englische wurde mit „guter Unterstützung“ bewertet, gefolgt von Sprachen wie Niederländisch, Französisch, Deutsch, Italienisch und Spanisch mit „moderater Unterstützung“. Sprachen wie Baskisch, Bulgarisch, Katalanisch, Griechisch, Ungarisch und Polnisch zeigen nur „fragmentarische Unterstützung“, womit auch sie in die Gruppe von Hochrisiko-Sprachen eingeordnet werden müssen.

„Die Ergebnisse unserer Studie sind äußerst alarmierend. Der Mehrheit der europäischen Sprachen mangelt es in gravierendem Maße an technologischer Unterstützung, und einige werden fast komplett vernachlässigt. In diesem Sinne sind viele unserer Sprachen noch nicht zukunftssicher“, sagt Prof. Hans Uszkoreit, Koordinator von META-NET, wissenschaftlicher Direktor am DFKI und zusammen mit Dr. Georg Rehm Mitherausgeber der Studie. Dr. Georg Rehm (DFKI) fügt hinzu: „Es gibt dramatische Unterschiede bei der Unterstützung der verschiedenen europäischen Sprachen und Technologiebereiche durch Sprachtechnologie. Die Schere zwischen „großen“ und „kleinen“ Sprachen geht immer weiter auf. Wir müssen sicherstellen, dass wir alle kleineren und unterversorgten Sprachen mit den nötigen Basistechnologien ausstatten, denn sonst sind diese Sprachen zum digitalen Aussterben verdammt.“

Sprachtechnologie ist Software, die Sprache – in geschriebener oder gesprochener Form – verarbeitet oder generiert. Bekannte Beispiele für Anwendungen, die sprachtechnologische Komponenten enthalten, sind Textverarbeitungen und Mailprogramme (Rechtschreibkontrolle, Grammatiküberprüfung), automatische Übersetzungsprogramme, persönliche Assistenten in Smartphones (z.B. Siri im iPhone), automatische Dialogsysteme für Auskünfte oder Buchungen per Telefon, Suchmaschinen im Web und synthetische Stimmen in Navigationssystemen. Derzeit basieren sprachtechnologische Komponenten und Systeme vornehmlich auf statistischen Methoden, die immens große Datenmengen geschriebener oder gesprochener Sprache benötigen. Gerade für Sprachen mit vergleichs-

Prof. Dr. Annette Schavan, Bundesministerin für Bildung und Forschung: „Europas Mehrsprachigkeit und unsere wissenschaftliche Kompetenz prädestinieren dazu, die Sprachtechnologie als informationstechnologische Herausforderung voranzutreiben. META-NET eröffnet Wege für die Entwicklung überall einsatzfähiger multilingualer Sprachtechnologie.“

weise wenigen Sprechern ist es oftmals sehr schwierig, diese Datenmengen im benötigten Umfang zu sammeln. Zudem unterliegen statistische Verfahren inhärenten Grenzen, was ihre Qualität betrifft, wie man z.B. an den oftmals amüsanten fehlerhaften Übersetzungen sehen kann, die von Online-Übersetzungsprogrammen erzeugt werden.

Europa hat es geschafft, fast alle Grenzen zwischen seinen Staaten abzubauen. Jedoch existieren immer noch die scheinbar undurchdringlichen, unsichtbaren Sprachbarrieren, die den freien Fluss von Wissen und Informationen behindern. Dies beeinträchtigt auf lange Sicht auch das Ziel, einen einzigen digitalen Markt zu schaffen, da Waren, Produkte und Dienstleistungen nicht ungehindert fließen können. Sprachtechnologie hat das Potenzial, diese Sprachgrenzen mit Hilfe moderner maschineller Übersetzungsmethoden zu überwinden. Die META-NET-Studie zeigt jedoch klar, dass viele europäische Sprachen dafür noch nicht hinreichend ausgestattet sind. Die technologischen Lücken sind signifikant und entstehen vor allem durch die Fokussierung auf die englische Sprache in Forschung und Entwicklung, mangelndes Engagement und geringe finanzielle Ressourcen sowie nicht zuletzt durch fehlende klare Forschungs- und Technologievisionen.

Um die fehlenden Basistechnologien für die Mehrheit der Sprachen zu entwickeln und den Technologietransfer anzustoßen, bedarf es umfassender Anstrengungen und einer präzisen Koordination. Es gibt stichhaltige Gründe, diese große Herausforderung mit einer gemeinsamen Strategie zu adressieren, welche die Europäische Union, ihre Mitgliedsstaaten und die Industrie einbezieht. Diese Gründe sind unter anderem die hohe finanzielle Pro-Kopf-Belastung kleinerer Sprachgemeinschaften, der notwendige Technologietransfer zwischen den Sprachen, die fehlende Interoperabilität der Ressourcen, Werkzeuge und Dienste sowie der Umstand, dass Sprachgrenzen oft nicht mit politischen Grenzen übereinstimmen.

### **Sprachtechnologie: Hintergrund**

Sprachtechnologie unterstützt uns schon heute bei alltäglichen Aufgaben, beispielsweise dem Schreiben von Emails oder dem Buchen von Flügen. Ohne uns dessen bewusst zu sein, profitieren wir von Sprachtechnologie, wenn wir Webseiten suchen oder übersetzen, die Rechtschreib- oder Grammatikkontrolle einer Textverarbeitung benutzen, den Bordcomputer im Auto oder unser Smartphone mit Sprachbefehlen steuern, in einem Online-Shop Empfehlungen bekommen oder den gesprochenen Anweisungen eines Navigationssystems folgen. Schon bald werden wir uns sowohl mit Computerprogrammen als auch mit Maschinen und Haushaltsgeräten unterhalten können. Wo auch immer wir uns aufhalten werden, wenn wir eine Information benötigen – wir werden einfach danach fragen, sobald wir Hilfe brauchen. Computer und Smartphone werden in der Lage sein, die Frage zu analysieren und bestmöglich zu reagieren. Der Wegfall der Kommunikationsbarriere zwischen Mensch und Maschine wird unsere Welt verändern.

Seit etwa 2011 wird Sprachtechnologie als eines der zentralen Wachstumsgebiete innerhalb der Informationstechnologie betrachtet. Große internationale Unternehmen wie Google, Microsoft, Apple, IBM und Nuance haben beachtliche Summen in die Entwicklung sprachtechnologischer Anwendungen wie z.B. Siri (Apple) oder Watson (IBM) investiert. In Europa gibt es tausende kleiner und mittlerer Unternehmen, die teils hochgradig spezialisierte sprachtechnologische Produkte oder Sprachdienstleistungen anbieten. Sprachtechnologie erlaubt es Menschen, zusammen zu lernen, zusammen zu arbeiten und Wissen über Sprachgrenzen hinweg zu teilen – und all das unabhängig von ihren Computerkenntnissen.

### **Die META-NET White Paper Serie**

Die META-NET Language White Paper Serie „Europe’s Languages in the Digital Age“ berichtet über den Stand von 30 europäischen Sprachen in Bezug auf Sprachtechnologie und erklärt die dringendsten Risiken und Chancen. Die Serie deckt alle offiziellen Sprachen der EU-Mitgliedstaaten und verschiedene andere in Europa gesprochene Sprachen ab. Zwar existiert eine Reihe von grundlegenden, umfassenden wissenschaftlichen Studien über bestimmte Aspekte einzelner Sprachen und Technologien, doch gab es bis jetzt kein allgemein verständliches Kompendium, das die wichtigsten Erkenntnisse und Herausforderungen für jede Sprache im Hinblick auf ein technologie-unterstütztes mehrsprachiges Europa darstellt. Diese Lücke füllt die META-NET White Paper Serie. META-NET präzisiert, warum die meisten Sprachen ernstzunehmenden Problemen gegenüber stehen, und zeigt zielgenau die bedrohlichsten Lücken auf. Insgesamt trugen mehr als 200 Experten aus Wissenschaft und Industrie zur Erstellung der White Papers bei.

Die White Papers haben eine Länge von je ca. 100 Seiten und wurden für die folgenden europäischen Sprachen geschrieben: Baskisch, Bulgarisch, Katalanisch, Kroatisch, Tschechisch, Dänisch, Niederländisch, Englisch, Estnisch, Finnisch, Französisch, Galizisch, Deutsch, Griechisch, Isländisch,

Irish, Italienisch, Lettisch, Litauisch, Maltesisch, Norwegisch, Polnisch, Portugiesisch, Rumänisch, Serbisch, Slowakisch, Slowenisch, Spanisch, Tschechisch und Ungarisch.

Jedes White Paper wurde in der Sprache verfasst, über die es berichtet und beinhaltet eine vollständige englische Übersetzung.

## Über META-NET und META

META-NET ist ein Exzellenz-Netzwerk, das aus 60 Forschungseinrichtungen in 34 Ländern besteht und das Ziel hat, eine technologische Basis für die mehrsprachige Europäische Informationsgesellschaft aufzubauen. META-NET wird flankiert von META, der offenen Technologie-Allianz für das mehrsprachige Europa. Mehr als 600 Organisationen (Unternehmen und wissenschaftliche Einrichtungen) in 55 Ländern sind bereits Mitglieder dieses Bündnisses für Sprachtechnologien.

## Weitere Informationen

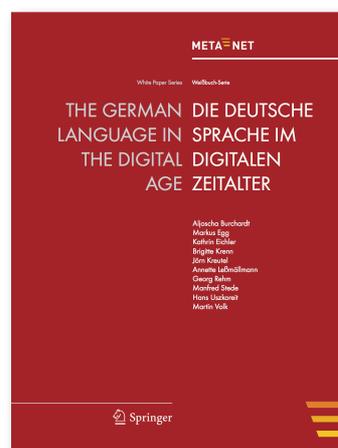
<http://www.meta-net.eu>  
<http://www.meta-net.eu/whitepapers>  
<http://www.meta-net.eu/whitepapers/press-release>  
<http://www.meta-net.eu/whitepapers/volumes/german>

META-NET  
White Paper Serie  
Pressemitteilung in ca. 30 Übersetzungen  
„Die deutsche Sprache im digitalen Zeitalter“

## Kontakt

Prof. Hans Uszkoreit und Dr. Georg Rehm  
META-NET c/o DFKI GmbH  
Alt-Moabit 91c  
D-10559 Berlin

E-Mail: [georg.rehm@dfki.de](mailto:georg.rehm@dfki.de)  
Telefon: +49 30 23895-1833  
<http://www.meta-net.eu>



„Die deutsche Sprache im digitalen Zeitalter“  
<http://www.meta-net.eu/whitepapers/e-book/german.pdf>

Das META-NET Exzellenz-Netzwerk wird u.a. finanziert über das 7. Rahmenprogramm und das ICT-PSP-Programm der Europäischen Kommission durch die folgenden Finanzhilfvereinbarungen: T4ME Net (Nr. 249119), CESAR (Nr. 271022), META-NET4U (Nr. 270893) und META-NORD (Nr. 270899).