

Überwindung der Sprachbarrieren

Automatische Sprachübersetzung und Untertitelung in der Anwendung

Nach dem Wegfall von Handelsbarrieren und Reisebeschränkungen innerhalb der europäischen Union verbleibt die Sprachbarriere in Europa als größtes Hindernis, das einem wahrhaftig vereinheitlichten europäischen Kultur- und Wirtschaftsraum entgegensteht. Die Autoren betrachten die Einsatzmöglichkeiten von Systemen zur automatischen Sprachübersetzung und Untertitelung anhand verschiedener Anwendungsbeispiele aus den Bereichen TV, Web und Hochschule.

Ob persönliche Begegnungen oder der Zugriff auf elektronische Inhalte über das Internet: in vielen Bereichen wäre es nötig, Inhalte von einer Sprache in andere Sprachen zu übersetzen, damit alle Europäer sie verstehen können – so etwa Fernsehsendungen. Technisch ist es zwar möglich, Sendungen aus allen europäischen Ländern zu empfangen, doch das Fehlen der notwendigen Sprachkenntnis versperrt großen Teilen potentieller Zuschauer den Zugriff.

Die Mengen der Inhalte, die übersetzt werden müssten, ist dabei so groß, dass nicht genügend menschliche Expertise zur Verfügung steht, um sie manuell übersetzen zu können. Auch wären die Kosten für menschliche Übersetzungen in vielen Fällen prohibitiv hoch, so dass selbst bei ausreichend zur Verfügung stehenden Übersetzern, diese aus Kostengründen nicht eingesetzt werden würden. Ähnliche Probleme entstehen bei der monolingualen Verschriftung von multimedialen Inhalten. Eine Verschriftung allen vorhandenen Audio- und audiovisuellen Materials ist aus Kosten- und Ressourcengründen nicht möglich.

Techniken zur automatischen Verarbeitung von Sprache, wie etwa automatische Spracherkennung oder Sprachübersetzung, bieten hier eine mögliche Lösung, um diese Lücken zu füllen.

Projekt EU-BRIDGE

Sprachverarbeitungstechniken wie Spracherkennung oder maschinelle Übersetzung sind immer noch wenig robust, und führen bei leichten Fehlbedienungen zu prohibitiv schlechten Ergebnissen. Auch liefern sie nur dann eine ausreichend gute Leistung, wenn sie auf die jeweilige Erkennungs- und Übersetzungsaufgabe maßgeschneidert sind. Beides, die Bedienung der Techniken und deren Anpassung auf spezifische Aufgaben,

erfordert hochqualifizierte Experten. Diese Experten sind aber häufig bei den Entwicklern von Anwendungen, die Sprachtechniken bei sich integrieren wollen, nicht vorhanden. Existierende APIs sind häufig zu komplex für Anwendungsentwickler aus anderen Bereichen; existierende Sprachverarbeitungssysteme sind häufig allgemein gehalten und lassen sich vom Anwender schwer oder gar nicht auf neue Aufgaben anpassen.

Das von der Europäischen Union geförderte Projekt Bridges Across the Language Divide (EU-BRIDGE) ist angetreten, um diese Probleme zu beheben. Durch die Einführung einer, wenn gewünscht, netzwerkbasiereten Service-Architektur mit einer einfach zu bedienenden Schnittstelle soll die Brücke zwischen Sprachtechnikern und Anwendungsentwicklern geschlagen werden. Auf Seite der Benutzer der Service-Architektur wird dabei nur eine Internetanbindung benötigt, die über genügen Bandbreite zum Versand der zu bearbeitenden Daten verfügt. Für die Anbindung existieren Beispielkunden für die gängigen Betriebssysteme. Auf der Seite des Dienstansbieters werden die Anfragen in einem Linuxcluster verarbeitet, indem ggf. die notwendigen Ressourcen dynamisch allokiert werden. Bei der automatischen Spracherkennung werden dabei insbesondere leistungsstarke CPUs benötigt oder im optimalen Fall sogar GPUs, während der notwendige Speicherverbrauch sich dadurch im Rahmen hält, dass mehrere Prozesse sich speicherintensive Komponenten teilen können. Für die maschinelle Übersetzung sind sehr große Mengen an Hauptspeicher von Vorteil um leistungsstarke Modelle anzuwenden zu können. Dies erhöht die Leistungsfähigkeit und die Laufgeschwindigkeit der Übersetzungsmodelle.

Die Vision von EU-BRIDGE ist es, Sprachtechnik für solche Anwendungsbereiche zu entwickeln, in denen sie bislang nicht einge-

setzt werden konnten. Das Projekt konzentriert sich dabei insbesondere auf die Techniken der automatischen Spracherkennung und der maschinellen Übersetzung. Es betreibt dabei nicht nur Forschung in den Grundlagen der jeweiligen Techniken, sondern strebt eine zügige Markteinführung an. Um den Nutzen der Service-Architektur von EU-BRIDGE zu demonstrieren, entwickelt das Projekt verschiedene Anwendungsbeispiele, in denen die neuen Technologien konkret und auf dem Markt zum Einsatz kommen. Innerhalb des Projekts entwickelten die zehn Partner aus Forschung und Industrie gemeinsam automatische Verschriftungs- und Übersetzungsangebote verschiedener Sprachen. Konkret wurden folgende Szenarien entwickelt:

- Automatische Untertitelung von Fernsehsendungen
- Automatische simultane Übersetzung von Universitätsvorlesungen
- Sprachübersetzungsservice für das Europäische Parlament
- Automatische Untertitelung und Übersetzung von Webinaren

Die entwickelten Technologien stehen Unternehmen zur Verfügung, die auf einem multilingualen, audio-visuellen Markt agieren und

Dr. Sebastian Stüker leitet die Nachwuchsgruppe „Multilinguale Spracherkennung“ am Karlsruher Institut für Technologie. Er

forscht seit mehr als zehn Jahren im Bereich der automatischen Spracherkennung und Sprachübersetzung. Im Rahmen des Projektes EU-BRIDGE befasst er sich mit der automatischen Untertitelung von Fernsehsendungen und der simultanen Übersetzung von Vorlesungen.



Prof. Dr. Alexander Waibel ist Professor der Informatik am Karlsruher Institut für Technologie und an der Carnegie

Mellon University, Pittsburgh, PA, USA. Er ist Direktor des internationalen Center for Advanced Communication Technologies (interACT), seine Forschungsschwerpunkte sind Spracherkennung, Sprachverarbeitung, Sprachübersetzung, multimodale und perzeptionelle Benutzerschnittstellen, sowie neuronale Netze. Alexander Waibel ist Koordinator vieler internationaler Projekte unter anderem von EU-BRIDGE.



diesen zu einem Vorteil in einer globalisier-
ten, multilingualen Welt verhelfen.

Verfügbare Sprachen

Es gibt etwa 6.000 verschiedene Sprachen auf der Welt, dazu kommen noch die unterschiedlichen Dialekte. Das Projekt hat seinen Fokus auf diejenigen europäischen und außereuropäischen Sprachen gelegt, die für die Europäer von besonderer Bedeutung sind (Bild 1). Das schnelle Übertragen der Technologien von einer Sprache auf die andere ist ein weiterer Teil der Forschung im Projekt. Vor allem wendet es sich der Anwendbarkeit auf unterversorgte Sprachen („Under-Resourced Languages“) zu.

Grundlegende Technik

Zwei Technikbereiche sind von besonderer Bedeutung innerhalb des Projekts: automatische Spracherkennung (ASR) und maschinelle Übersetzung (MT).

Spracherkennungssoftware hat dabei die Aufgabe, Aufnahmen menschlicher Sprache in Wortsequenzen in maschinenverarbeitbarer Form umzuwandeln. Wichtig ist dabei zu beachten, dass gesprochene Sprache, die, wie von der ASR, Wort für Wort niedergeschrieben wird, sich stark von den gewohnten Texten, die wir kennen, unterscheidet. Gesprochene Sprache ist in der Regel ein Wasserfall von Worten, der, anders als geschriebene Sprache, nicht strukturiert ist, etwa durch Satzzeichen oder Absätze. Auch ist gesprochene Sprache häufig sehr ungrammatisch und fragmentarisch und daher in niedergeschriebener Form schwer zu verstehen. Deshalb bedarf es weiterer Hilfstechniken, um das Ergebnis der Spracherkennung aufzubereiten, z. B. mit Satzzeichen anzureichern,

Bearbeitete Sprachen in EU-BRIDGE

Sprache X	MT von		
	ASR	Eng. von X	X nach Eng.
Englisch	X	-----	-----
Französisch	X	X	X
Spanisch	X	X	X
Deutsch	X	X	X
Italienisch	X	X	X
Chinesisch		X	X
Polnisch	X	X	X
Slowenisch		X	X
Portugiesisch	X	X	X
Russisch	X		X
Arabisch	X	X	X
Türkisch	X		

Bild 1. Übersicht der verfügbaren Sprachen

von Hesitationen und Fragmenten zu bereinigen oder weiter zu strukturieren.

Maschinelle Übersetzung übersetzt geschriebenen Text von einer Sprache in geschriebenen Text in einer anderen Sprache. Sie erwartet dabei als Eingabe wohl strukturierten Text mit Satzzeichen und Segmentierung in Sätze. Wenn man jetzt Spracherkennung und maschinelle Übersetzung in Reihe schaltet, um Sprachübersetzungssysteme zu erhalten, muss man spezielle Vorkehrungen an der Schnittstelle beider Komponenten treffen, um die Wortfolge, die man aus der Spracherkennung erhält, für die maschinelle Übersetzung aufzubereiten. Neben Bereinigung der ASR-Ausgabe, Segmentierung in Sätze, Anreicherung mit Satzzeichen, muss man sich auch um solche Dinge kümmern, wie die Wortfolgen gesprochener Zahlen in die zugehörige Ziffernfolge umzuwandeln

oder Akronyme korrekt aufzubereiten. Auch sollten die Vokabularien der beiden Systeme aufeinander abgestimmt werden, so dass die maschinelle Übersetzung in der Lage ist, auch alle Sätze, die die Spracherkennung erkennen kann, zu übersetzen.

Spracherkennung und maschinelle Übersetzung sind stochastische Systeme, die ihre Modelle auf Trainingsdaten lernen. Für Spracherkennung benötigt man große Mengen (am besten mehrere hunderte Stunden) auf Satzebene transkribierter Audiodaten in der Zieldomäne. Ferner benötigt man große Menge an Texten (mehrere Millionen bis zu Milliarden laufender Wörter), diesmal ohne Audiodaten, die gut auf die Erkennungsaufgabe passen.

Zum Training maschineller Übersetzungssysteme benötigt man ebenfalls große Mengen (Hundert Tausende bis Millionen) übersetzter Sätze, sowie sehr große Mengen an monolingualen Text in der Zielsprache.

Spracherkennungs- und Übersetzungssysteme funktionieren am besten, wenn sie möglichst genau auf die Zieldomäne zugeschnitten sind. Spracherkennungssysteme können z.B. nur Wörter aus einem vorher definierten Vokabular erkennen, und sind besonders genau, wenn sie auf Sprecher und akustische Umgebung trainiert sind. Die manuelle Anpassung solcher Systeme ist zeitaufwendig und kostenintensiv. Deshalb arbeitet EU-BRIDGE an Techniken zur automatischen und autonomen Anpassung der Systeme und zur kostengünstigen Akquisition von Adaptionsdaten.

Vier Anwendungen als Ideengeber

Vier Beispiele zeigen, dass die entwickelten Technologien in der Praxis und in der Anwendung einsatzfähig sind (Bild 2). Sie sind als

bic4

- Automation
- Digitale Archivierung
- Migration auf HD
- Workflow Optimierung

www.bic4.de

ADVANCED SYSTEM INTEGRATION

EURO MEDIA
SERVICE

OTT-Encoding • Event Stream Management
OTT-Monitoring • CDN Service (Live & VoD)

Tel.: +49 2421 69 36 051 - www.euromedia-service.de

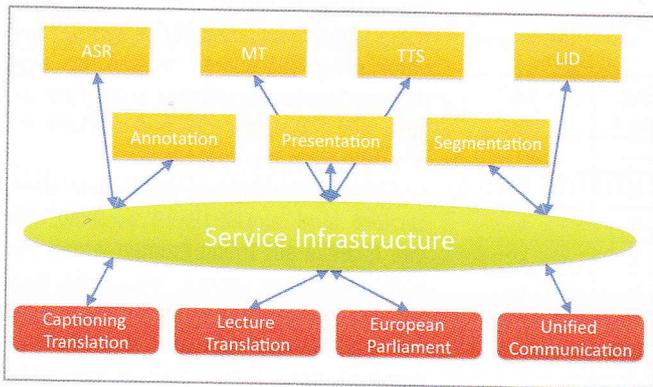


Bild 2. Übersicht Servicearchitektur



Bild 3. Lecture Translator: Seit dem Jahr 2012 werden ausgewählte Vorlesungen am Karlsruher Institut für Technologie automatisch simultan übersetzt.

Ideengeber gedacht und sind auf zahlreiche weitere Anwendungsbereiche übertragbar:

Automatische Untertitelung von Fernsehsendungen

Der Inhalt von Fernsehsendungen muss für hörgeschädigte Personen in Untertiteln wiedergegeben werden. Auch wenn dieser Untertitelungsprozess heute in hohem Maße optimiert ist, so ist er dennoch sehr kostspielig und aufwändig im Produktionsprozess. Grund dafür ist zum einen der hohe Anteil an manueller Arbeit sowie die sehr spezifische Ausbildung der Mitarbeiter. Untertitel werden häufig durch das sogenannte „Re-speaking“ erstellt: Der Untertitler hört das Audio des Fernsehprogramms und wiederholt es, frei von Hintergrund- und Hallgeräuschen, in eine personalisierte Spracherkennungssoftware. Da Untertitel häufig live erstellt werden müssen, spielen enge Zeitvorgaben eine entscheidende Rolle und es besteht in der Regel keine Möglichkeit, Fehler zu korrigieren. Maschinelle Unterstützung kann die anstrengende Arbeit der menschlichen Untertitler in verschiedenen Bereichen unterstützen:

Automatische Untertitelung von Wettervorhersagen

Wettervorhersagen haben einen sehr spezifischen Inhalt. Ein neu entwickeltes System macht es nun möglich, eine genauere Verschriftung zu liefern als bisherige sprecherunabhängige Spracherkennungsprogramme. Bedingt durch das hohe Genauigkeitsniveau sind nur sehr geringe menschliche Korrekturen nötig, bevor es ausgestrahlt werden kann. Projektpartner Red Bee Media Limited hat in Testläufen mit Wettervorhersagen festgestellt, dass das System einen Mehrwert bietet.

Mehrsprachige Untertitelung von Nachrichten

Nachrichtensendungen müssen teilweise in verschiedene Sprachen übersetzt werden. Dafür wird im ersten Schritt mithilfe einer spezifischen Spracherkennungssoftware die Sprache in Text umgewandelt und anschließend für ausgewählte Sprachen maschinell übersetzt. Nachrichtensendungen des Senders Euronews können mit den Systemen bereits untitelt und teilweise übersetzt werden.

Für beide Anwendungsfälle kommen spezielle, automatische Spracherkennungssysteme zum Einsatz, die auf die jeweiligen Fernsehsendungen zugeschnitten wurden. Die Ausgabe der Erkennungssysteme wird danach entsprechend aufbereitet, in Sätze segmentiert und mit Satzzeichen versehen.

Projektpartner

Akademiker, Ingenieure und Wirtschaftsexperten aus zehn Forschungseinrichtungen und der Industrie entwerfen gemeinsam kompatible Lösungen für die verschiedenen Anwendungsbereiche der Übersetzungstechnologien. Partner sind:

- Karlsruher Institut für Technologie, Deutschland
- Fondazione Bruno Kessler, Italien
- Polish Japanese Institute of Information Technology, Polen
- RWTH Aachen, Deutschland
- The University of Edinburgh, Vereinigtes Königreich
- The Hong Kong University of Science and Technology, China
- Red Bee Media Limited, Vereinigtes Königreich
- Pervoice SpA, Italien
- Accipio Projects GmbH, Deutschland
- Andrexen, Frankreich

Wichtig ist dabei, dass die Formatierungsvorgaben und Nebenbedingungen, die für die Untertitel der jeweiligen Sender gelten, eingehalten werden. Die Anpassung der Spracherkennungssysteme für die Wettervorhersagen ist dabei auf Archiven von Sendungen mit bereits erstellten Untertiteln erfolgt. Dadurch konnten die Kosten für Daten niedrig gehalten werden. Allerdings war das Training auf Untertiteln eine Herausforderung, da die Zeitmarken der Untertitel nicht genau genug waren. Ferner stellen Untertitel nicht immer eine wortwörtliche Verschriftung des Gesagten dar, so dass traditionelle Trainingsmethoden suboptimal funktionieren und neue Trainingstechniken entwickelt wurden.

Die Spracherkennungssysteme für Euronews wurden sogar nur auf der Audiospur alter Ausstrahlungen trainiert, ohne dass irgendeine Verschriftung vorlag. Einzig die Webseiten des Senders gaben zusätzliche Informationen.

Automatische simultane Übersetzung von Universitätsvorlesungen

Die Internationalisierung stellt auch Universitäten vor sprachliche Herausforderungen: Sollen sie ihre Lehrveranstaltungen auf Englisch umstellen, oder müssen alle ausländischen Studierenden Deutsch lernen? Eine alternative Lösung ist der an einer Universität weltweit erste automatische simultane Übersetzungsdienst. Der Vorlesungsübersetzer zeichnet automatisch den Vortrag des Referenten auf, verschriftet diese Sprachsignale und übersetzt sie in Echtzeit ins Englische. Studierende können über Laptop, Smartphone oder Tablet der Vorlesung folgen. Zudem übersetzt die Technologie auch die Vorlesungsfolien und ermöglicht den Zugriff auf vergangene Veranstaltungen über Suchbegriffe in den verschrifteten Vorträgen. Von dem Service profitieren nicht nur ausländische

sche, sondern auch hörgeschädigte Studierende. Die Übersetzung ist nicht immer perfekt, ermöglicht aber den Studierenden, der Vorlesung deutlich besser zu folgen. Das System ist am Karlsruher Institut für Technologie (KIT) seit dem Jahr 2012 in der Anwendung. (Bild 3).

Auf technischer Seite stellt der Vorlesungsübersetzer eine Zusammenschaltung von Spracherkennung und maschineller Übersetzung dar. Von besonderer technischer Herausforderung ist dabei die Erkennung in Echtzeit sowie die möglichst schnelle Segmentierung der Ausgabe in übersetzbare Einheiten, damit die Übersetzung mit möglichst geringer Latenz angezeigt werden kann. Das Übersetzungssystem adaptiert sich auch mit der Zeit an die jeweiligen Vorlesungen, indem es unüberwacht auf den Aufnahmen der Vorlesenden lernt und Zusatzmaterialien, wie Vorlesungsfolien und Webseiten, auswertet.

Sprachübersetzungsservice für das Europäische Parlament

Das Europäische Parlament arbeitet in seinen 24 Amtssprachen, diese entsprechen

insgesamt 552 möglichen Kombinationen, da jede Sprache in 23 andere übersetzt werden kann. Hier können automatische Technologien die Dolmetscher und Übersetzer des Europäischen Parlaments unterstützen. Eine webbasierte Anwendung hilft bei der Suche nach Fachausdrücken und bei der Identifizierung von Eigennamen. Testversionen des Services sind beim Europäischen Parlament in der Erprobung.

Automatische Untertitelung und Übersetzung von Webinaren

Webinare werden häufig mit Teilnehmern aus verschiedenen Ländern und mit unterschiedlichen Sprachen gehalten. Dabei stellt die Sprachbarriere oftmals ein Hindernis dar. In diesem Anwendungsfall wird die Sprachübersetzungstechnik, die auch beim Vorlesungsübersetzer zum Einsatz kommt, in die Webinarplattform „Serenty“ integriert. Die Plattform ermöglicht, Webinare über das Netz sowie reguläres Telefon abzuhalten. Sie läuft als Webanwendung in modernen Browsern. Das Audio des Vortragenden kann dabei entweder über den Browser aufgenommen wer-

den, oder aber auch über eine reguläre Telefonverbindung, per Einwahl in den Serenty-Server. Über die Browserapplikation sieht jeder Teilnehmer den Videoinhalt des Seminars oder auch die Folien des Vortragenden. Ferner können die Seminarteilnehmer sich auch über die Webapplikation per Audi zu Wort melden oder sich alternativ per Telefon einwählen. Zusätzlich steht die Möglichkeit zur Versendung von Soforttextnachrichten zur Verfügung. Die Einbindung des Übersetzungssystems geschieht dabei auch über die Soforttextfunktion. Das Übersetzungssystem agiert als virtueller Teilnehmer, der kontinuierlich die Übersetzung des Seminars als Soforttext verschickt.

Ausblick

Die entwickelten maschinellen Übersetzungstechnologien sind angewandte Beispiele, die Sprachbarriere zu überbrücken und zeitgleich die Sprachvielfalt aufrecht zu erhalten. Sie sind Grundlage für eine Vielzahl weiterer Anwendungsszenarien im Bereich der automatischen multilingualen Spracherkennung und Übersetzung. |

Beschleunigung und Automatisierung Globaler Content Workflows

Egal wo sich ihr Content befindet: in der Cloud, am eigenen Standort oder am anderen Ende der Welt, Aspera Software ermöglicht Zugriff, Übertragung, Austausch, Weiterverarbeitung, Verwaltung, Synchronisierung, Zustellung und Verbreitung mit maximaler Geschwindigkeit - und das ohne Limits.

Asperas einzigartige, patentierte FASP™ Transporttechnologie ist der Motor für highspeed Datenübertragung und Automatisierung um weltweit komplexe Medien-Workflows zu beschleunigen.

Erfahren Sie mehr über unsere Datentransfer-Lösungen und überzeugen Sie sich selbst wie Aspera Ihr Unternehmen effektiv verändern kann!

 **aspera**
an IBM® company