

Kurze Einführung in die PROMT Machine Translation Technologie

Was ist MT? Historischer Hintergrund

Die maschinelle Übersetzung ist eine Technologie, die Prozesse bereitstellt, die das automatische Übersetzen von Texten jeder Art aus einer natürlichen Sprache in eine andere Sprache ohne menschliche Intervention möglich machen. *Machine Translation Technology* (MT) basiert auf tiefgehende Forschungsergebnisse auf dem Gebiet der Linguistik und des *Natural Language Processing* sowie auf der Anwendung innovativer Computertechnologien bei der Übersetzung von Texten aus einer natürlichen Sprache in eine andere Sprache. Als eine der frühesten Weiterentwicklungen in der Computerinformatik hat sich MT als ein schwer erreichbares Ziel erwiesen. Heute steht eine Vielzahl von Systemen zur Verfügung, deren Übersetzungsergebnisse, wenn auch nicht perfekt, dennoch in einigen Spezialbereichen von einer ausreichenden Qualität und nützlich sind.

Die Anfänge der Entwicklung von Systemen für die automatische Übersetzung liegen 50 Jahre zurück – die Anwendung dieser Systeme bei der Übersetzung von Fremdsprachen war faktisch bereits seit der Erfindung des elektronischen Computers in den 40er Jahren Gegenstand der Forschung. Viele Jahre lang basierten diese Systeme primär auf der direkten Übersetzung mittels zweisprachiger Wörterbücher mit einer relativ wenig detaillierten Analyse der Syntax-Struktur. In den 80er Jahren jedoch machten Fortschritte in der rechenbetonten Linguistik die Anwendung anspruchsvollerer Methoden möglich, und in einigen Systemen wurde bei der Ausführung von Übersetzungsaufgaben eine indirekte Methode angewandt. In diesen Systemen werden die Texte der Eingabesprache unter Einbezug von aufeinander folgenden Programmen für die Identifikation von Wortstrukturen (Morphologie) und Satzstrukturen (Syntax) und für die Lösung des Problems der Mehrdeutigkeit als abstrakte Darstellungen der „Bedeutung“ analysiert.

Die abstrakten Darstellungen müssen eindeutig sein und die Basis für das Generieren von Texten in einer Zielsprache oder in mehreren Zielsprachen zur Verfügung stellen. Es gab in der Tat zwei „indirekte“ Grundmethoden. Bei der einen Verfahrensweise, die auf dem „*interlingua*“-Prinzip basiert, ist die abstrakte Darstellung so konstruiert, um eine Art sprachenunabhängige „*interlingua*“ zu sein, die potentiell als Vermittler zwischen einer Vielzahl von natürlichen Sprachen fungieren kann. Das Übersetzen findet in zwei Grundschritten statt: aus der Eingabesprache in die *interlingua* – und aus der *interlingua* in die Ausgabesprache. Beim zweiten indirekten Lösungsansatz, der allgemein bekannter ist und auf dem Prinzip des Transfers basiert, wird die Darstellung zunächst in eine äquivalente Darstellung für die Ausgabesprache konvertiert. Folglich gibt es hier drei Grundschritte: Analyse des Eingabetextes in einer abstrakten Ursprungsdarstellung, Transfer in eine abstrakte Zieldarstellung und das Generieren in der Ausgabesprache.

Systeme dieser Art wurden seit den späten 80er Jahren entwickelt. Und heute noch können alle gegenwärtig verfügbaren und angebotenen Systeme in drei Grundsystemtypen eingeteilt werden: direkt, interlingual, „Transfer“. Die bekanntesten MT-Systeme für Großrechner zählen eigentlich zum Transfer-Übersetzungstyp. Bei den PROMT-Systemen werden Übersetzungen nach der Transfermethode ausgeführt. Darüber hinaus sind PROMT-Übersetzungssysteme verbesserte Versionen dieses Typs. Im Gegensatz zu ihren Vorgängern weisen sie einen komplex modularen Aufbau auf und sind einfach zu verändern und zu erweitern.

PROMT wurde 1991 gegründet. Während der letzten 13 Jahre haben hoch qualifizierte Linguisten und Programmierer in enger Zusammenarbeit einzigartige linguistische Technologien und Übersetzer-Software entwickelt, die in der Industrie der maschinellen Übersetzung weltweit konkurrenzlos sind.

Gegenwärtig entwickelt PROMT eine große Auswahl an Tools für die maschinelle Übersetzung, deren Einsatz in allen Umgebungen auf PCs, Mac, im Internet oder Intranet möglich ist, entweder als Standalone-Anwendungen (z. B. Textverarbeitung) oder als Übersetzungsfunktionen, die direkt in eine Anwendung (Intranet, Internet, Word oder Excel) eingebunden werden.

Die Vorteile von MT

Die wichtigsten Vorteile der Technologie der maschinellen Übersetzung sind:

Hohe Geschwindigkeit. Das maschinelle Übersetzen kann die Zeit, die notwendig ist, um große Textmengen zu übersetzen, erheblich reduzieren.

Kostenreduktion. Durch die Reduzierung der Zeit, die für routinemäßige und wiederkehrende Übersetzungsprozesse aufgewandt wird, kann der Einsatz von MT zu einer Einsparung von bis zu 50 % der gesamten Übersetzungskosten führen.

Konsistenz der Terminologie. Die Verwendung von umfangreichen Wörterbüchern, effizient organisierten Datenbanken und vorprogrammierten linguistischen Algorithmen, um die Informationen über die Eingabesprache zu analysieren, gewährleistet eine signifikante Erhöhung der Übersetzungskonsistenz.

Übersetzungs-Anforderungen – Wofür wird MT verwendet?

Beim Versuch, einen allgemeinen Überblick über die Entwicklung und Verwendung von MT-Systemen und Übersetzungstools zu geben, ist es wichtig, zwischen sechs Grundtypen von Übersetzungs-Anforderungen zu unterscheiden.

Das Wesentliche. Da MT-Systeme keine qualitativ hochwertigen Übersetzungen liefern konnten (und es immer noch nicht können), haben Anwender herausgefunden, dass sie das, was sie wissen wollten, dem unüberarbeiteten Ausgabertext entnehmen konnten. Sie waren eher mit einem spärlich übersetzten Text zufrieden, als gar keine Übersetzung zu haben. Mit dem wachsenden Angebot billigerer PC-basierter Systeme wuchs auch diese Art und Weise der Anwendung schnell und beträchtlich.

MT ist eine Komponente der Informationszugriffs-Systeme. Integration von Übersetzungssoftware in folgende Datenbanken: (1) Systeme für das Suchen und Wiederauffinden von ganzen Dokumenttexten in Datenbanken (im Allgemeinen elektronische Versionen von Zeitschriftenartikeln aus der Wissenschaft, Medizin und Technologie) oder für das Wiederfinden von bibliografischen Informationen; (2) Systeme für den Auszug von Informationen (z. B. Produktdetails) aus Texten, vor allem aus Zeitungsberichten; (3) Systeme für die Textzusammenfassung; (4) Systeme für die Recherche in nicht-textuellen Datenbanken stellen dem Anwender einen bevorzugten Zugriff auf Informationsquellen zur Verfügung.

MT für die Lokalisierung. Unternehmen stehen heute vor der Herausforderung, global zu kommunizieren: mit Kunden, Verkäufern, Partnern und insbesondere mit den Mitarbeitern, die alle in verschiedenen geografischen Regionen agieren. MT wird benutzt, um Firmeninformationen,

PROMT®

16 Birzhevaya liniya, Vasilyevsky Ostrov,

St. Petersburg, 199034 Russia

tel.: +7 (812) 331-7540, fax: +7 (812) 327-44-83

info@e-prompt.com

www.prompt.com

www.online-translator.com

die in einer Sprache verfasst wurden, Sprechern der anderen Sprachen zuzustellen. Produkte des Unternehmens werden in den neuen Märkten lokalisiert. Die exzellente Geschwindigkeit von MT-Systemen bei der Übersetzung umfangreicher routinemäßiger und wiederkehrender Schriften macht diese Systeme zu einzigartigen Lokalisierungs-Tools.

MT für die Kommunikation. Die Anforderung an die Übersetzungen von Texten im Internet wie Websites, E-Mails und sogar Chat-Listen wachsen mit großer Geschwindigkeit. In diesem Kontext kommt die Möglichkeit einer herkömmlichen Übersetzung nicht in Betracht. Die Notwendigkeit einer zeitnahen Übersetzung ist unabdingbar, um den Grundinhalt der Mitteilungen zu übermitteln, egal wie spärlich die Information ist.

MT für professionelle Zwecke. MT kann zur Produktivität und Effizienz der Arbeit professioneller Übersetzer beitragen: Spezialtools für die Terminologiesuche und ein umfangreicher Satz von individuellen Anwender-Einstellungen, die in den Systemen der maschinellen Übersetzung integriert sind, erweisen sich als effektive Werkzeuge für professionelle Übersetzer.

MT als Teil der automatisierten Übersetzungsprozesse. Die Integration der Machine Translation Technology in die Struktur der *Translation Memory* Systeme (TM) eröffnet eine wichtige Perspektive für die Entwicklung einer voll automatisierten technologischen Prozesskette, die effektiv angepasst werden kann, um sehr komplizierte Übersetzungsaufgaben zu bewältigen.

PROMT Technology: Wie funktioniert sie?

Der übersetzte Text muss in der Zielsprache so richtig wie nur möglich sein und die Information, die im Originaltext enthalten ist, wiedergeben – das ist die wichtigste Herausforderung, mit der die Entwickler von MT-Systemen heute konfrontiert werden. Um eine qualitativ hohe und inhaltsgetreue Übersetzung zu liefern, werden die Anbieter von Technologien der maschinellen Übersetzung folgende Aufgaben in Angriff nehmen müssen:

(1) *Wörterbücher* sind die größten Komponenten eines MT-Systems, was die Menge der Informationen, die sie umfassen, betrifft. Wenn sie mehr als nur Wortlisten sind (und sie müssten es sein, damit das System gut funktioniert), dann könnten sie sehr gut die teuersten Komponenten sein, die es zu entwickeln gilt. Die Wörterbücher der MT-Systeme beinhalten nicht nur die Übersetzung von Wörtern und Kollokationen, sondern auch spezielle morphologische, grammatikalische und semantische Attribute/Merkmale, die bei der Analyse und Synthese verwendet werden.

(2) *Übersetzungsmodule*, die komplizierte Strukturen linguistischer Algorithmen beinhalten, die die morphologischen und grammatikalischen Strukturen der Ein- und Ausgabesprache beschreiben.

(3) *Eine Reihe von zusätzlichen Software-Tools* für die Erstellung und Bearbeitung von Anwender- und Fachwörterbüchern, Terminologie-Tools, einen linguistischen Editor, Einstellungstools, benutzerfreundliche Oberflächen, Tools für die Textnachbearbeitung.

Aufgrund der einmaligen Wörterbucharchitektur und der innovativen linguistischen Algorithmen, die in allen PROMT-Produkten integriert sind, bieten PROMT-Systeme für die maschinelle Übersetzung für alle diese Probleme effektive Lösungen an.

Multidimensionale Wörterbuch-Architektur. Die Wörterbücher in den existierenden MT-Systemen unterscheiden sich im Format, Inhalt, Detail-Level und präzisen Formalismus für die lexikalische

Beschreibung. Unterschiedliche Theorien der linguistischen Darstellung können zu unterschiedlichen Betrachtungen der Wörterbücher führen, und unterschiedliche Arten von MT-engines stellen unterschiedliche Anforderungen an die Inhalte des Wörterbuchs.

Weil PROMT-Systeme die Übersetzung nach der Transfer-Methode durchführen, liefern die Wörterbücher der PROMT-Systeme Informationen über die Begriffe der Eingabesprache und über ihre Übersetzungen, über die Begriffe der Ausgabesprache und über ihre Übersetzungen. Das System benutzt eine formale Methode der morphologischen Beschreibung als Grundlage für das Wiederauffinden von Wörterbucheinheiten. In den PROMT-Übersetzungssystemen ist die morphologische Beschreibung für alle Übersetzungsrichtungen entwickelt worden. Diese Beschreibung ist fast einzigartig in ihrer Größe. Sie umfasst 800 Flexionsarten für die russische Sprache, mehr als 300 Flexionsarten für Deutsch und Französisch und 250 Flexionsarten für Englisch. Die Vielfalt an Flexionsformen für jede Sprache ist als Baumstruktur gespeichert, so dass sowohl eine effektive Modalität der Informationsspeicherung als auch eines effektiven Algorithmus der morphologischen Analyse gewährleistet sind.

Diese signifikante Optimierung der Hauptwort-Struktur im Wörterbucheintrag reduziert redundante und sich wiederholende Informationen im Wörterbuch: Es ist nicht länger erforderlich, separate Wörterbucheinträge für alle stammverwandten Beugeformen (Konjugationen, Deklinationen usw.) zu machen, weil diese Art von Informationen jetzt in einem Einzeleintrag gespeichert sind. Das morphologische Modell der lexikalischen Informationsspeicherung wurde auch für die Entwicklung des Beratungs-Systems für die Anwender integriert, die ihre persönlichen Wörterbücher innerhalb der System-Struktur erstellen. Wenn der Anwender einen neuen Wörterbucheintrag macht, entnimmt das System automatisch den Wortstamm und weist ihm die passende Flexionsform zu.

Ein neuer Mechanismus für das Speichern und die Assoziation der Wörterbucheinträge, der in @prompt 7.0 implementiert ist, bietet weitere Möglichkeiten für die Verbesserung der Wörterbucharchitektur. Die innovative multidimensionale Methode wurde für die Struktur der Wortbeschreibung eingeführt: Jedem Wort oder jeder Redewendung wird mindestens eine aktive Übersetzung zugeordnet. Darüber hinaus können sie mehreren inaktiven Übersetzungen innerhalb einer Redewendung zugeordnet werden. Aktive Übersetzungsvarianten werden im direkten Übersetzungsprozess verwendet, während die nicht-aktiven Übersetzungsvarianten nach zusätzlichen Informationen über die Bedeutung eines Wortes durchsucht werden können.

Jede nicht-aktive Übersetzungsvariante kann in eine aktive konvertiert werden und umgekehrt. Die multidimensionale Wörterbucharchitektur macht es möglich, eine unbegrenzte Anzahl von Übersetzungen aus elektronischen und/oder gedruckten Wörterbüchern einzugeben, um die Zeit und den Aufwand bei der Erstellung und Verwaltung von Wörterbucheinträgen zu reduzieren und um mit einer unbegrenzten Anzahl von Übersetzungsvarianten zu arbeiten.

PROMT Linguistik Algorithmen. Im Gegensatz zur allgemeinen linguistischen Verfahrensweise, die die Implementierung von aufeinander folgenden Prozessen der Satzanalyse und -synthese voraussetzt, basiert die Architektur der PROMT-Systeme auf der Darstellung von Übersetzungsprozessen als „objektorientierte“ Prozesse. Diese Methode berücksichtigt die Anwendung verschiedener Algorithmen auf unterschiedlichen Übersetzungsebenen.

Die Übersetzungs-Algorithmen der PROMT-Systeme basieren auf einer hierarchischen Verfahrensweise, die eine Unterteilung von Übersetzungsprozessen in miteinander verbundenen Prozessen für die verschiedenen Einheiten der linguistischen Analyse voraussetzt. Man unterscheidet folgende Levels im System: lexikalisches Level, Gruppen-Level, Level des einfachen Satzes und Level des zusammengesetzten Satzes. Alle diese Prozesse sind miteinander verbunden und beeinflussen sich gegenseitig hierarchisch in Übereinstimmung mit der Hierarchie

der Texteinheit und tauschen synthetisierte und geerbte Attribute aus. Diese Art der Algorithmus-Anordnung erlaubt die Anwendung verschiedener formaler Methoden für die Beschreibung der Algorithmen auf den verschiedenen Levels.

Die lexikalische Beschreibung einer Einheit innerhalb eines Wörterbucheintrags ist eng verbunden mit der Struktur der System-Algorithmen und ist nicht auf der Basis von „syntaktisch-semanticen“ Gegensätzen konfiguriert, sondern eher auf der Basis von Wortkomponenten-Levels. Demnach können PROMT-Systeme unter Verwendung von unvollständig beschriebenen Wörterbucheinträgen arbeiten, was sehr wichtig ist, wenn Wörterbücher für Anwender erstellt werden, die mit der Methode der linguistischen Analyse nicht vertraut sind.

PROMT Anpassungstools – wie kann die Qualität von MT verbessert werden?

MT-Systeme arbeiten mit natürlichen Sprachen – einem Datensatz, der unendlich variiert werden kann, mehrdeutig und strukturell komplex ist. Um adäquat zu übersetzen, muss ein MT-System die Kenntnis von Hunderten von Syntax-Strukturen, Variationen und Ausnahmen sowie Beziehungen zwischen diesen Strukturen kodieren. Software für die maschinelle Übersetzung muss mit einem immer wechselnden Wortschatz und mit spezifisch semantischem Wissen über den Sprachgebrauch von Sprachmustern von zigtausend von Wörtern angeboten werden. Das System muss die präzise Identifikation der Wortarten und der grammatikalischen Charakteristiken der Wörter sicherstellen, die in verschiedenen Kontexten Substantive, Verben oder Adjektive sein können, so dass jedes Wort viele möglichen Übersetzungen haben kann. Übersetzen macht ein riesiges Speichern an Wissen über die Welt, die Kommunikationsabsichten und über den Inhalt erforderlich.

Es gibt sie immer noch: die Meinung, dass MT noch nicht „gut genug“ für die kommerzielle Nutzung ist. Befürworter der MT brauchen umfangreiche, leicht zu bedienende Tools, um die Qualität ihrer Übersetzungen zu verbessern. Die Tools müssen für Nicht-Entwickler, die aber die Sprachen und die Business-Terminologie ihres Unternehmens kennen, zugänglich sein. Unter der Handvoll kommerzieller MT-Systeme, die heute angeboten werden, ist PROMT ein Anbieter, der das Problem der Qualität effektiv und erfolgreich gelöst hat. Wir haben Anpassungstools entwickelt, die die Qualität der Übersetzung signifikant verbessern. PROMT-Produkte kennzeichnen folgende Eigenschaften, die zu einer hochqualitativen Übersetzung beitragen.

Themenvorlagen werden benutzt, um mehrere Wörterbücher, Präprozessoren und reservierte Wörter (die nicht übersetzt werden können/sollen) nach Themenbereichen zu kategorisieren und ihre Priorität zu bestimmen. Themenvorlagen beinhalten einen Satz anwenderspezifischer Einstellungen, die dem Themenbereich entsprechend erstellt wurden und anpassbar sind. Der Anwender kann Themenvorlagen speichern und sie beim Übersetzen von Texten aus spezifischen Themenbereichen benutzen – dies kann die Qualität der Übersetzung signifikant erhöhen, die Konsistenz bei der Verwendung spezieller Begriffe gewährleisten und den Übersetzungsprozess effektiver machen. Bei der System-Installation der Software wird automatisch eine allgemeine Themenvorlage mit der Bezeichnung „Allgemein“ erstellt. Nach der Installation zusätzlicher Fachwörterbücher stehen dem Anwender auch andere Themenvorlagen zur Verfügung. Die Themenvorlagen können angepasst werden, der Anwender kann auch andere Kategorien von Themenvorlagen erstellen, wenn Fachwörterbücher installiert sind.

Die Integration des *Dictionary Editor* in den PROMT-Produkten stellt den Anwendern Optionen für die Bearbeitung der Wörterbücher zur Verfügung, die einzigartig in ihrem Umfang und in ihrer Anpassungsfähigkeit sind: Anwender können Informationen in den Anwenderwörterbüchern

PROMT®

16 Birzhevaya liniya, Vasilyevsky Ostrov,

St. Petersburg, 199034 Russia

tel.: +7 (812) 331-7540, fax: +7 (812) 327-44-83

info@e-prompt.com

www.prompt.com

www.online-translator.com

einsehen, erstellen und editieren. Sie können mit jedem ausgewählten Wort/jeder ausgewählten Wortverbindung arbeiten. Sie können mit Wörterbucheinträgen der Quell- und Zielsprache in einem Wörterbuch oder in mehreren Wörterbüchern arbeiten, erweiterte grammatikalische und morphologische Informationen einsehen und bearbeiten, bidirektionale Wörterbücher miteinander verknüpfen, Wörter aus Textdokumenten einfügen oder importieren. Bei der Erstellung neuer Wörterbucheinträge können Anwender sowohl das Übersetzungsäquivalent selbst als auch grammatikalische Informationen zum Wort eintragen, einschließlich dessen Flexionsart und Rektion, die eine korrekte Übersetzung des Wortes gewährleisten. Der Dictionary Editor stellt den Anwendern auch die Möglichkeit bereit, eine Liste mit Wörtern zu speichern, die nicht übersetzt werden müssen – Die Liste der reservierten Wörter beinhaltet Firmennamen und Produktnamen, die transliteriert werden können, um die Konsistenz der Firmen-Warenzeichen und -Politik zu gewährleisten.

Fachwörterbücher können installiert werden, um die Qualität der Übersetzung von Texten aus speziellen Themenbereichen zu verbessern. Alle Fachwörterbücher enthalten vordefinierte Themenvorlagen.

Der *Associated Memory Manager* wird verwendet, um die Eingabetexteile und ihre äquivalenten Übersetzungen direkt in die AM-Datenbank für eine spätere Verwendung zu speichern. Der Anwender kann AM-Übersetzungsdatenbanken erstellen, löschen oder manuell editieren, nach Datenbankeinträgen suchen und diese ersetzen.

Mit den professionellen Tools für PROMT-Wörterbücher können Anwender deren Status checken, statistische Informationen abfragen und Wörterbucheigenschaften einsehen oder editieren, Wörterbücher der gleichen Übersetzungsrichtung miteinander vergleichen und verknüpfen.

PROMT *Terminology Manager* (PROMT TerM) stellt den Anwendern ein automatisiertes Tool für die Terminologie-Extraktion und -Pflege sowie für das -Management zur Verfügung.

Synopsis

PROMT (www.prompt.de) ist ein weltweit führender Anbieter von Technologien für die maschinelle Übersetzung. Das Unternehmen wurde 1991 gegründet, und seine Sachkenntnis umfasst über 13 Jahre Entwicklung von Technologien für die maschinelle Übersetzung. Durch die Verbindung von Innovation und technischer Perfektion hat PROMT bahnbrechende Technologien für die maschinelle Übersetzung (MT) entwickelt und wurde zum technologischen Führer in der MT-Industrie. Die Kerntechnologie von PROMT strebt anwenderspezifische Übersetzungslösungen an, die tiefreichendes linguistisches Wissen sowie technologisch und gestalterisch führende Computertechnologien in sich vereinen.

Heute bietet PROMT eine große Auswahl an Übersetzungslösungen an: Systeme für die maschinelle Übersetzung und Services, Wörterbücher, Translation Memory Systeme, Datenpflege-Systeme für alle Business-Bereiche und Spezialisten sowie für das Internet/Intranet. Unsere Unternehmenspolitik der kontinuierlichen Innovation in den Bereichen Software und Zustellungsservice hat einen wachsenden Marktanteil in all diesen Bereichen erreicht. Die neue Software-Produktlinie für Übersetzungen @prompt umfasst eine große Auswahl an hoch entwickelten Anwendungen und Lösungen, die die von PROMT in den 13 Jahren erzielte Qualitätssteigerung weiterführen. Die @prompt Übersetzungs-Software wird als die höchst entwickelte auf dem Markt gewertet und garantiert die höchst mögliche Qualität und Effizienz der Übersetzungsprozesse.

Unsere Kunden schätzen unser Engagement in der sich fortentwickelnden Forschung und unsere Investitionen auf dem Gebiet der Linguistik und des *Natural Language Processing*: Die Spannweite unserer Auftraggeber und Kunden reicht von SME bis hin zu auf dem Weltmarkt präsenste Unternehmen, einschließlich international führender Unternehmen wie Xerox, SAP, Lukoil, General Motors, Volvo Car International, Deloitte & Touch u. a.

PROMT`s Auftrag ist es, mit der Entwicklung einer neuen Generation von MT-Systemen, die Übersetzungen von hoher Qualität gewährleisten, eine barrierefreie weltweite Geschäftstätigkeiten in möglichst vielen Bereichen zu ermöglichen.