

Die Weltübersetzung: Web to TM

Durch die Globalisierung ist das Übersetzungsvolumen explodiert. Selbst KMU mit wenigen Mitarbeitenden drängen über das Internet auf den Weltmarkt, so dass immer mehr übersetzt werden muss. Die maschinelle Übersetzung liefert keine überzeugenden Resultate, weswegen die Zahl der Übersetzer ebenfalls explodieren müsste. Dem ist aber nicht so! Wie kommt es, dass immer mehr Seiten von wenigen Übersetzern bewältigt werden können?

Recycling per Fuzzy-Logik

Die Erfindung des Translation Memory (TM) hat die seit babylonischen Zeiten stillstehende Übersetzungstradition, deren unverzichtbarer Bestandteil dicke Wörterbücher waren, grundlegend aus den Fugen gebracht. Translation Memorys nutzen den Umstand aus, dass Menschen zu repetitiven Verhaltensweisen neigen. Viele Firmen und Organisationen führen über Jahre hinweg ähnliche Prozesse durch, die sich aus linguistischer Sicht gleichen wie ein Zebra dem anderen. Dieses sogenannte Firmenwording – aus statistischer Sicht betrachtet immer das gleiche Unternehmenssprech – kann mithilfe der TMs optimal eingefangen werden. Die kleinen Wunderwerke arbeiten mit Fuzzylogik, also einer unscharfen Logik, die in der Lage ist, variable Sprachmuster zu erkennen. Wenn ein Firmenchef alle Jahre eine ähnliche Weihnachtsrede hält, lassen sich die Übersetzungskosten von Jahr zu Jahr senken.

Die ersten kommerziellen Übersetzungsspeicher stammten aus Stuttgart. Zwei Nerds, Jochen Hummel und Iko Knyphausen, gründeten 1984 eine kleine Übersetzungsagentur und begannen in den späten 1980er Jahren mit der Entwicklung von Übersetzungshilfsprogrammen, auch CAT-Tools (Computer Aided Translation) genannt. Das System funktioniert einfach: Einzelne übersetzte Textelemente werden zusammen mit dem Originaltext in einer Datenbank abgelegt und bilden so ein Parallelkorpus. Falls Trados, so heisst die mittlerweile weltbekannte Software, einen Textbaustein oder Teile davon wiedererkennt, recycelt sie die Übersetzung aus dem Speicher. Im Idealfall muss der Übersetzer den vorgeschlagenen Textbaustein nur noch validieren, so dass seine Arbeit schneller voranschreitet. Dank CAT-Tools können bei repetitiven Texten signifikante Einsparungen erzielt werden. Wenn die Automarke Citroen die Handbücher für den C2, C3, C4, C5 produziert hat, entstehen beim Handbuch für den C8 viel weniger Kosten.

Der überwältigende Vorteil von TMs ist unsichtbarer Natur. TMs standardisieren die Sprache. Sie standardisieren die Dokumentation, den Verkauf, den Support und den Internetauftritt. Das gesamte Übersetzungsmanagement vereinfacht und vereinheitlicht sich. Die erste global erfolgreiche Vereinheitlichungsorgie wurde 1997 gefeiert, als Microsoft sein Betriebssystem möglichst einfach in alle wichtigen Weltsprachen übersetzen wollte. Bill Gates kaufte kurzerhand einen Teil von Trados auf, puschte das System und machte aus Windows die erfolgreichste Software aller Zeiten. Es folgten Dell und andere Grosskunden. 2005 kaufte die britische Konkurrenzfirma SDL das Unternehmen auf und übernahm Trados als Markennamen für ihre Softwareprodukte.

Von der Insellösung zum Cloud TM

Weitaus der grösste Teil der Übersetzer arbeitet als Freelancer bei sich zuhause. Schnelle Datenleitungen waren bis vor kurzem eher die Ausnahme. In der Not legte jeder Übersetzer seine eigene Datenbank an, wodurch eine Unzahl von Insellösungen entstand. Die Übersetzer betrachteten ihre Speicher als Privatbesitz, weswegen es schwierig war, sie zur Zusammenarbeit zu motivieren. Die grossen Agenturen konterten mit Zwangskollektivierung. Sie schufen einen neuen Beruf: den Translation Manager. Wie ein Steuereintreiber forderte dieser nach jedem Auftrag die zweisprachigen Dateien zurück. Die Übersetzer und Korrektoren bekamen fortan Auftragspakete, die alle wichtigen Übersetzungsressourcen enthielten. Das Arbeitsresultat musste als Retourpaket geschickt und danach vom Translation Manager in den Hauptspeicher eingelesen werden.

Smarte Agenturen konnten enorme Speicher anlegen und ihre Marktposition ausbauen. Ihr Fleiss wurde allerdings durch einen systeminhärenten Fehler getrübt. Je grösser der Hauptspeicher, desto schlechter geriet die Qualität der Pakete. Der sogenannte Konkordanzspeicher und die mittlerweile gigantischen Wörterbücher passten nicht mehr ins

Paket, was den einfachen Empfänger gegenüber dem Benutzer eines Servers stark benachteiligte.

2015 brachte die Firma SDL, welche die Trados-Software übernommen hatte, den ersten brauchbaren GroupShare-Server auf den Markt. Die Textbausteine werden seitdem auf einem zentralen Server abgespeichert, so dass beliebig viele Übersetzer auf die Übersetzungsressourcen zugreifen können. Dank der Server-Technologie sind dem Recycling von Übersetzungsbausteinen keine Grenzen mehr gesetzt. Bis zur Weltübersetzung, einem Cloud-TM, das alle übersetzten Segmente des Webs enthält, ist es nur noch ein kleiner Schritt.

Die digitale Revolution als Kostenkiller

Das erste vollautomatische Übersetzungsnetz wurde 1987 in Genf noch vor dem Web unter dem Namen TTN Translation Network konzipiert. Der Gründer, Martin Bächtold, probierte im Silicon Valley an der Stanford University die ersten interuniversitären Netzwerke aus. In den Vorlesungen erklärten die Dozenten das Modell des komparativen Kostenvorteils, und sofort war klar: Übersetzung und Kommunikation gehen fortan Hand in Hand. Übersetzungen würden künftig am billigsten und qualitativ besten Standort produziert, also in einer Region, wo die Zielsprache aktiv gesprochen wird.

Bei der Rückkehr nach Genf lag eines der ersten Modems im Gepäck. Mithilfe dieser damals in der Schweiz noch verbotenen und laut zischenden Blechbox wurde auf einem Schnyder PC mit einer 10-Mega-Festplatte der erste Übersetzungsserver der Welt installiert. Doch diese Innovation kam viel zu früh für den Markt. Kein Mensch wusste damals, wie ein Modem funktionierte. Die Firma musste Geld aufnehmen, um in Taiwan günstig Geräte einzukaufen, die sie gratis an Kunden und Übersetzer verschickte. Zu den ersten Kunden zählte der Lawinenwarndienst vom SLF in Davos. Die Lawinenwarnungen mussten extrem schnell übersetzt, sodann der Text nicht über Fax, sondern im digitalen Format übermittelt werden. Wenn ein Lawinenbulletin eintraf, avisierten laute Faxwarnungen die Übersetzer, ein Ursystem, an dessen Stelle heute längst SMS und Smartphone-Interface getreten sind.

1989 wurde das Worldwide Web im Genfer CERN geboren und revolutionierte durch einen neuen Standard die Kommunikationstechnologie. Mit Kundennummer 16 bei der damaligen Post startete

TTN ins Internet. Den Gewinn aus dem ersten System investierte man in die Entwicklung einer Art Arpa-Netz für Übersetzungen in Indien, wo ein riesiges IT-Team den Code programmierte. Anhand eines replizierten Netzes sollte ein Cloud-System geschaffen werden, das 165 Sprachen vollautomatisch routen kann. Der Versuch endete kläglich, der Code war zu lang, die Probleme viel komplexer als angenommen.

Der zweite Anlauf war erfolgreicher. Es dauerte aber viel länger als erwartet. Schritt um Schritt konnten immer grössere Teile der Prozesse automatisiert und die Produktionspreise um 30% gesenkt werden. Es zeichnet sich ab, dass Agenturen, die künstliche Intelligenz einsetzen, grosse Kundenportfolios effizienter verwalten als ihre rein menschlichen Pendanten. Die Programme berechnen die Auslastung der Übersetzer unter Berücksichtigung der Arbeitszeiten und Ferienabwesenheiten. Dank dem optimierten Zeitmanagement profitieren die Übersetzer von einem konstanteren Arbeitsfluss. Es gibt weniger Stress bei erhöhter Produktivität.



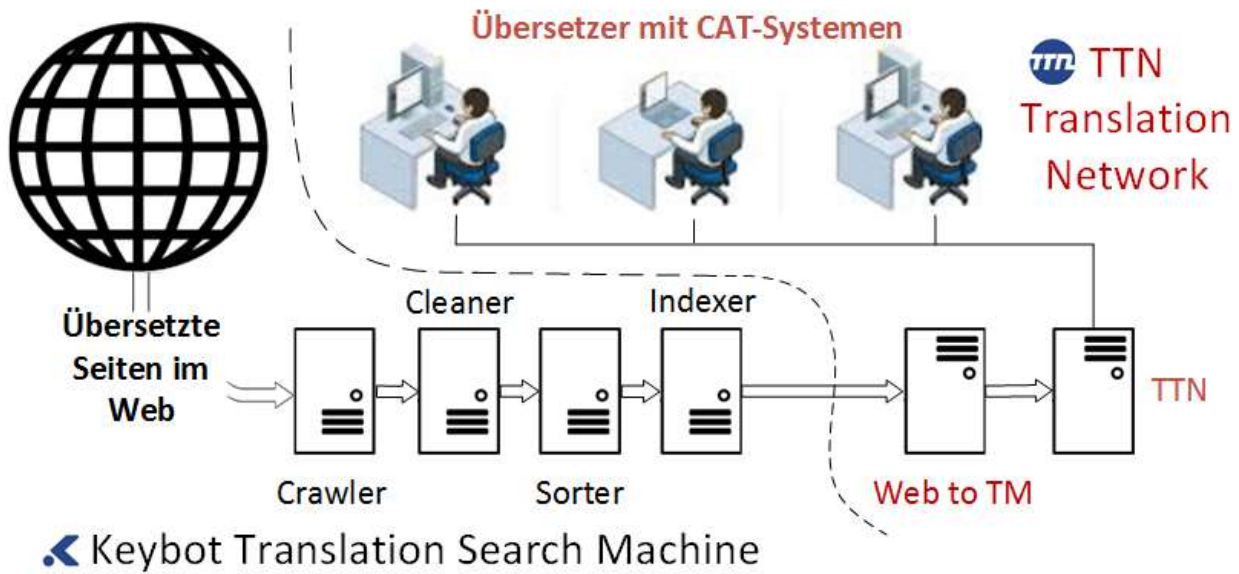
Hohe Rechenleistung gefragt!

Patrick Boulmier von Info-Logo parametrisiert zusammen mit Martin Bächtold, CEO Keybot, die neuesten Hochleistungsrechner für die Weltsprachmaschine. Pro Minute müssen mehrere Hundert Webseiten in TMs konvertiert werden.

Keybot: Web to TM

Bei vielen weltweiten Konzernen verläuft die Digitalisierung im Übersetzungsbereich chaotisch. Sie verfügen über Webapplikationen mit abertausenden übersetzten Seiten, haben aber keine Übersetzungsspeicher, wo diese Texte sauberlich in Parallelkorpora abgelegt sind. Die Nachlässigkeit bei der Auswahl des Übersetzungsproviders hat verheerende Folgen: Wenn schlechtorganisierte Firmen ihren Webauftritt überarbeiten möchten, müssen sie für jede Seite den vollen Preis bezahlen, weil die bereits geleistete Arbeit nicht recycelt werden kann. Viel Wissen geht unnötig verloren, und die Wiederbeschaffung kostet Geld.

Web to TM: Das Web wird abgesaugt und in ein gigantisches Translation Memory (TM) umgewandelt



Web to TM soll solchen Firmen wieder auf die Sprünge helfen. Die Firma Keybot, eine Tochterfirma von TTN, entwickelte die gleichnamige Übersetzungssuchmaschine, die wie Google das Web durchforstet. Sie speichert ausschliesslich mehrsprachige Seiten und indexiert diese als Parallelkorpora. Ein komplexes Netz von Servern betreibt Data-Mining, indem es die Internetseiten potentieller Kunden nach übersetzten Textelementen absucht. Das abgesaugte Wissen, also die Big Data, muss gereinigt, sortiert und statistisch ausgewertet, Wiederholungen gezählt, ihre Signifikanz berechnet und abgespeichert werden. Erst wenn dieser aufwendige Prozess abgeschlossen ist, kann die Maschine Häppchen um Häppchen an eine Batterie von GroupShare-Servern senden. Wenn dann nach diesem langen Prozedere ein Übersetzer einen Auftrag mit seiner CAT-Software öffnet, übersetzt sie automatisch alle Teile, die die Suchmaschine auf der Webapplikation des Kunden gefunden hat. Der Übersetzer verfügt immer über die neueste, publizierte Version und nicht eine veraltete Fassung, die firmenintern nachbearbeitet wurde.

Damit Keybot die Sprachelemente zuordnen kann, liest sie alle Wikipedia-Seiten in 165 Sprachen sowie die Übersetzungen der Bibeltexte und der Menschenrechte ein. Jede Sprache besitzt ihren eigenen Gencode, der in Form von N-Grammen extrahiert werden kann. Diese statistischen Eigenschaften sucht Keybot zu nutzen, um die Textbausteine zu identifizieren und zu parallelisieren. Das

System befindet sich immer noch im Betastadium, und bis jetzt können nur zuverlässige TMs generiert werden, wenn ein Kunde seinen Webaufttritt so strukturiert hat, dass sich der Crawler beim Einlesen nicht verheddert. Der grösste bis jetzt generierte Übersetzungsspeicher wurde für eine amerikanische Firma hergestellt; er umfasst 23 Sprachen.

Keybot möchte das ganze mehrsprachige Web in einen gigantischen Übersetzungsspeicher umwandeln. **Web to TM** heisst die neue Devise. Die Transformation ist extrem rechenintensiv und kann nur von einer entsprechend grossen Server-Farm erledigt werden. Um das nötige Kapital zu beschaffen, bereitet Keybot in Deutschland den Börsengang an der KMU-Börse vor und versucht einen Teil des Maschinenparks via Crowdfunding zu finanzieren.

SLOTT Translation

Im Bereich der maschinellen Übersetzung kamen die entscheidenden Neuerungen aus der Wetterbranche. Warnmeldungen stecken in einem fast unlösbaren Dilemma: Einerseits müssen sie schnell verbreitet werden, andererseits dürfen sie keine Übersetzungsfehler enthalten. Der statistische Ansatz von Google Translate hilft bei dieser Aufgabe nicht weiter, weil er zu ungenau ist und nie in der Lage sein wird, die uhrwerkähnliche Präzision von Warnmeldungen wiederzugeben.

Der clevere Wetterfrosch Kachelmann, der in Zürich Mathematik studiert hatte, löste das Dilemma als erster. Er nahm ein einfaches Excel-Sheet und

bastelte ein System, mit dem er die Sprachgenerierung über Zellen steuern konnte. Bereits in den 80er Jahren versuchte der Direktor des Instituts für Schnee- und Lawinenforschung SLF, ein automatisches Übersetzungssystem zu bauen. Später scheiterte auch ein statistischer Versuch einer deutschen Uni, der mit dem Wahrscheinlichkeitsprinzip und Markow-Modellen arbeitete. Als dann Jahre später der Ingenieur Kurt Winkler vom SLF in Davos ein irre anmutendes Excel-Sheet von den Alpen in die Genfer Sprachmetropole schickte, lachten ihn Linguisten zuerst als «Fool on the hill» aus. Sein Projekt wanderte in die unterste Schublade wie ein schlechter Kriminalroman. Erst als er insistierte, beauftragte man einen mit Translation Memorys vertrauten Mitarbeiter mit der Falsifizierung. Ein falscher Satz, und Good bye! Winklers System wäre gestorben.

Nach drei Tagen herrschte immer noch Funkstille. Es wurden keine Fehler gefunden, selbst ein speziell zur Falsifizierung gebautes Programm vermochte keine nachzuweisen. Erstaunlich! Winkler, der von Linguistik keine Ahnung hat, wertete die Warntexte und deren Übersetzungen der letzten zehn Jahre nach Mutationsmöglichkeiten aus. Als Resultat entwickelte er eine Excel-Datenbank, die kein Mensch verstand.

Oder doch? Bereits vor hundert Jahren hatte der Genfer Ferdinand de Saussure, der Gründungsvater des Strukturalismus, in seinen Vorlesungen auf die Syntagmastruktur der Sprache hingewiesen. Als erster definierte er die Mutationsmöglichkeiten, die in einem Sprachgefüge vorkommen können, ohne jedoch die Verbindung zu anderen Sprachen herzustellen. Winkler zerlegte die Texte nach denselben Prinzipien und legte Transformationsregeln fest, anhand deren Textelemente von einer Sprache in eine andere übertragen werden können.



Maschinelle Übersetzungen für mehr Sicherheit

Dr. Kurt Winkler vom Institut für Schnee- und Lawinenforschung SLF gelang ein erstaunlicher Durchbruch im Bereich der automatischen Übersetzung. Lawinenwarnungen werden in einem Bruchteil einer Sekunde übersetzt.

Mithilfe von Winklers Satzkatalog lassen sich Abermillionen idiomatisch und grammatikalisch perfekter Sätze in vier Sprachen generieren. Das System funktioniert allerdings nur für Lawinenwarnungen in der Schweiz; die Sätze müssen über einen Katalog am Bildschirm generiert werden. Das ist nicht sehr praktisch und bietet nur beschränkte Anwendungsmöglichkeiten.

TTN experimentiert unter dem Namen SLOTT Translation mit einem ähnlichen System. Wie Wettermeldungen dürfen Übersetzungen keine Sprachfehler enthalten, da dies das Vertrauen der Kunden untergrübe. Anhand eines Katalogs von vorerst nur zwanzig Satzvorlagen will TTN in Zukunft die Kommunikation mit den Kunden standardisieren, damit Anfragen in allen Sprachen fehlerlos und korrekt beantwortet werden können.

Ob sich SLOTT als kommerzielles System behaupten kann, ist ungewiss. Ausser Zweifel steht, dass künftige TMs hierarchisch organisiert sein werden, weil dies ihr Potential erheblich vergrößert. Die nächste Generation von CAT-Systemen kann nicht nur einen im TM gespeicherten Text genau, sondern Abermillionen von Varianten korrekt übersetzen.

Wird der Übersetzer bald überflüssig?

TMs sind Recyclingmaschinen, die nicht kreativ denken können. Die aktuelle Prozessorgeneration ist von ihrer Architektur her in keiner Weise fähig, mit dem Synapsensystem des Menschen rivalisieren, wenn es darum geht, sinnvolle Sätze zu erzeugen. Der Übersetzer bleibt unersetzlich. Erst der Quantencomputer könnte diesen Zustand ändern.