

Star mit Millionen von Ventilatoren

Computerlinguistik der Universität Zürich im Dienst der Filmwirtschaft



SO NICHT Einfacher arbeitende Systeme produzieren oft Übersetzungen, über die man nur lachen kann. zvs

Die beiden Professoren am Institut für Computerlinguistik der Universität Zürich haben ein Startup-Unternehmen gegründet, das sich mit der maschinellen Übersetzung von Untertiteln für Filme befasst. Wenn alles gut geht, dürften sich bald auch Studierende daran beteiligen.

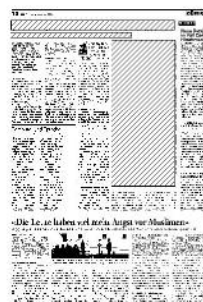
ALFRED BORTER

Das Institut für Computerlinguistik ist ein junges Pflänzchen, es entstand erst 2001, als Teil der Philosophischen Fakultät. Zuvor war die Computerlinguistik Teil des Instituts für Informatik. Geleitet wird das Institut seit seiner Gründung von Michael Hess; den zweiten Lehrstuhl hat Martin Volk inne.

Am Institut für Computerlinguistik sind etwa hundert Studierende eingeschrieben. In der Forschung befasst man sich am Institut unter anderem mit der automatischen Übersetzung von Texten, mit Systemen für die Extraktion genau spezifizierter Informationen aus Texten und mit dem Beantworten von Fragen über grosse Textmengen durch Computer.

Selbstlernendes System

Dieses Jahr haben die beiden



Professoren mit der TextShuttle GmbH ein Startup-Unternehmen gegründet; es wurde bereits für so zukunftssträftig befunden, dass es von der Swiss venture kick Initiative zwei Förderungsbeiträge bekommen hat.

TextShuttle befasst sich mit maschineller Übersetzung von Filmuntertiteln. Man steckt damit nicht mehr in den Kinderschuhen. Dank dem früheren Wirken von Professor Volk in Skandinavien arbeitet bereits eine auf die Übersetzung von Film-Untertiteln spezialisierte Firma mit dem auf computerlinguistischen Methoden basierenden System. Es erlaubt den Übersetzern, Filmuntertitel schneller, präziser und kostengünstiger in verschiedene Sprachen zu übertragen.



«Bei Seifenoper
 n ist das
 Programm
 besonders

effizient» MARTIN VOLK

Angefangen hat man mit der Übersetzung von Fernsehfilmen vom Schwedischen ins Dänische, dann auch ins Norwegi-

sche. Die Software schlägt automatisch eine Übersetzung vor, die nötigenfalls von einem professionellen Übersetzer noch verändert werden kann, wenn er mit dem Vorschlag nicht einverstanden ist. Das System ist selbstlernend, das heisst, dass es sich Eingriffe des Übersetzers «merkt» und bei einer nächsten Anwendung übernimmt. «Die Ergebnisse sind recht gut», hält Volk fest.

Krimis sind schwieriger

Übersetzt werden so in erster Linie Seifenoper, Kriminalfilme und selten auch Dokumentarfilme. «Bei Seifenoper», erklärt Martin Volk, «ist das Computerprogramm besonders effizient. Krimis sind schon schwieriger, und Dokumentarfilme sind ein echtes Problem.»

Warum? Weil in Seifenoper oft dieselben Aussagen vorkommen, etwa «Ich liebe dich» oder «Ich verlasse dich», während das in Krimis seltener ist und in Dokumentarfilmen kaum vorkommt.

Gegenwärtig ist die Marketing-Chefin der Firma daran, die Untertitelungsfirmen in ganz Europa auf die neue Dienstleistung aufmerksam zu machen.

«Wenn wir Erfolg haben, werden wir rasch auf die Mithilfe von Studierenden angewiesen sein», erklärt Michael Hess. Es sei für diese natürlich interessant, an Projekten mitzuwirken, die eine Chance haben, sich im Markt durchzusetzen.

Immer bessere Übersetzungen

Beide Professoren weisen darauf hin, dass die mithilfe von computerlinguistischen Methoden generierten Übersetzungen immer besser werden. Zum Glück, denn einfacher arbeitende Systeme produzieren oft Übersetzungen, über die man nur noch lachen kann. Etwa wenn im Bollywoodfilm «Om Shanti Om», der vor zwei Jahren an der Berlinale einen Riesenerfolg feierte, der Satz zu übersetzen war: «I know that you're a dream girl of millions of your fans.» Man las: «Ich weiss, dass Sie ein Traumädchen sind von den Millionen Ihrer Ventilatoren.» Nun kann das englische Wort fan tatsächlich Ventilator bedeuten, aber auch für einen fanatischen Verehrer stehen. Volk ist überzeugt: «Unser System würde sich für die richtige Version entscheiden.»

Computer und Sprache

Am Institut für Computerlinguistik kreuzten kürzlich Detmar Meurers von der Universität Tübingen und Stefan Riezler, bei Google Research in Zürich tätig, die Klängen: Gräbt die Sprachtechnologie der Computerlinguistik das Wasser ab? Für Riezler war klar: Wenn Milliarden oder gar Trillionen von Sprachdaten zur Verfügung stehen, genügt diese Menge, um zum Beispiel in der maschinellen

Übersetzung relevante Ergebnisse zu erhalten. Heute sei, das räumte er ein, die Qualität solcher Übersetzungen noch nicht über jeden Zweifel erhaben, aber sie verbessere sich ständig, weil dank der grossen Datenmenge eben auch seltene Aspekte genügend häufig erfasst würden, um Fehler auszuschliessen. Provokativ merkte er an, mit der Zeit werde die Computerlin-

guistik kaum mehr erforderlich sein. Meurers hingegen betonte, Daten allein machten noch keine gute Übersetzung, es brauche weiterhin Sprachwissenschaftler, welche dank Sprachanalyse zu besseren Ergebnissen kämen als reine Statistiker. Die Anwendung der Forschung im Bereich von Computerlinguistik und Sprachtechnologie geht längst über die Übersetzungsproblematik hinaus.

So erfuhr man, dass an gewissen nordamerikanischen Hochschulen die Essays, die von den Studierenden verlangt würden, nicht mehr ausschliesslich von Menschen begutachtet würden, sondern zusätzlich auch von Maschinen, die so programmiert sind, dass sie aufgrund der verwendeten Begriffe die Qualität der Essays recht zuverlässig bestimmen können. (ABR.)