

Maturandentag 2005

Computerlinguistik

<http://www.ifi.unizh.ch/cl/>

13./14. September 2005

Gebiet:

Eine relativ neue Disziplin

- ▶ im Überschneidungsbereich von *Linguistik* und *Informatik*
- ▶ mit dem Ziel, menschliche *Sprache* auf dem Computer zu *modellieren*

Fach:

Ab WiSe 06/07 in Zürich (unter “Bologna“) als

- ▶ Haupt- und Nebenfach
- ▶ je auf BA- und MA-Ebene (MA: ab WiSe 2008/09)

Sonst in der Schweiz:

- ▶ in Genf (z.Zt.) als Zusatzfach

Entwicklungsziele:

1. Entwicklung von *Anwendungsprogrammen*:

- ▶ Maschinelle Übersetzungssysteme
- ▶ Textzugriffs- und Textverwaltungssysteme
- ▶ natürlichsprachliche Schnittstellen zu Datenbanksystemen
- ▶ Systeme zum Generieren geschriebener Texte
- ▶ Systeme zum Verstehen gesprochener Sprache

2. Entwicklung von *Werkzeugen*:

- ▶ zur Corpusanalyse (Finden von Sprachmerkmalen in Texten)
- ▶ zum Aufbauen von Lexika (aus Texten, halb- oder vollautomatisch)
- ▶ zum Erstellen von Indizes (in Bibliotheken, für on-line Dokumentationen)
- ▶ zum Annotieren von Texten (Markieren bestimmter Ausdrücke)

Forschungsziele in Zürich:

- ▶ intelligenter Zugriff auf Dokumente
(v.a. Antwortextraktion und Informationsextraktion)
- ▶ (linguistisch intelligente) Systeme für das Fremdsprachenlernen
(kein automatisierter Drill, sondern Fehleranalyse)
- ▶ Systeme zum Auflösen von sprachlichen Mehrdeutigkeiten

Lehre in Zürich:

- ▶ Methoden für das Errechnen der grammatikalischen Strukturen von Sätzen
- ▶ Methoden für das Errechnen der Bedeutung von Sätzen
- ▶ Einsatz von sprachtechnologischen Werkzeugen in der Linguistik
- ▶ Vermittlung konkreter Programmierkenntnisse für die maschinelle Verarbeitung natürlicher Sprachen (keine Programmierkenntnisse vorausgesetzt)

Arbeitsbereiche/Branchen

- ▶ Universitäten
- ▶ Softwarehäuser, zum Beispiel für maschinelle Sprachverarbeitung, Datenbankentwicklung, sonstige Softwareentwicklung, Datenverarbeitungsdienste, Softwareberatung
- ▶ Verlagsgewerbe, insbesondere für Lexika, computerunterstützte Lehrmaterialien, neue Medien
- ▶ öffentliche Verwaltung, zum Beispiel nationale oder internationale Behörden
- ▶ Faxabrufservice, Sprachabrufservice
- ▶ Übersetzungsbüros, vor allem große Übersetzungsdienste
- ▶ Bibliotheken und Archive, insbesondere Dokumentationszentren und wissenschaftliche Bibliotheken

Aufgaben und Tätigkeiten

- ▶ Entwicklung und Optimierung von Software für maschinelle Sprachverarbeitung, inkl. Programme für maschinelle Übersetzung, Sprachgenerierung, Antwortextraktion etc.
- ▶ fachrelevante Daten (Dokumente, Literatur) ermitteln, dokumentieren und nutzbar machen
- ▶ ein- und mehrsprachige rechnergestützte Lexika und terminologische Wörterbücher erarbeiten
- ▶ Programme für rechnergestützten Unterricht bzw. computergestützte Selbstlernprogramme entwickeln
- ▶ Datenbanken und EDV-gestützte Archive aufbauen
- ▶ im Verlagswesen Manuskripte prüfen und bearbeiten, Autoren/Autorinnen betreuen, neue Produkte im Bereich Sprachwissenschaft/Computerlinguistik entwickeln und realisieren

Was ist denn so schwierig an der automatischen Verarbeitung der menschlichen Sprache?

Zum Beispiel die in der Sprache häufigen **Mehrdeutigkeiten**:

- ▶ Phonetisch: “catch it“ oder “cat sh...“?
- ▶ Lexikalisch: “Bank“, “Kiefer“, “Salat“
- ▶ Morphologisch
 - ▶ Wort- und Flexionsklassen: “can“
 - ▶ zusammengesetzte Wörter: “Schweinsschnitzel“ vs. “Jägerschnitzel“
- ▶ Syntaktisch: “Ich sehe den Mann im Park mit dem Feldstecher.“
- ▶ ... etc.

Computerlinguistik im Brennpunkt: Maschinelle Übersetzung

Demo: kommerziell gegen kostenlos

- ▶ kommerziell: “translate*pro*“
- ▶ kostenlos: Google’s Übersetzungsservice