

Studieninformationstage 2017

Computerlinguistik und Sprachtechnologie

<http://www.cl.uzh.ch/>

September 2017

- ▶ Was ist Computerlinguistik (CL): inkl. kleiner Demo
- ▶ Welche Berufsaussichten hat man?
- ▶ Wie kann CL an der Uni Zürich studiert werden?
- ▶ Voraussetzungen fürs Studium, Profil des CL-Studierenden

Zielsetzung

Computerlinguistik ist die automatische Verarbeitung der menschlichen Sprachen mit dem Computer.

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", "Berliner" (doghnut, citizen of Berlin)
- ▶ Morphologisch
 - ▶ Wort- und Flexionsklassen: "can"
 - ▶ zusammengesetzte Wörter: "Schweinsschnitzel" vs. "Jägerschnitzel"
 - ▶ zusammengesetzte Wörter: "Lärmschutzgesetz" vs. "Kinderschutzgesetz"
- ▶ Syntaktisch: "Ich sehe den Mann mit dem Feldstecher."
- ▶ ... etc.

Zwei Anwendungen: Demo

Suche im Internet

'Bank' .. ja schon, aber wer oder was genau ist gemeint?

Maschinelle Übersetzung: Google translate

(Es ist vollbracht:) Die Kiefer wurde von mir gefällt.

Koreferenzanalyse: <http://kitt.ifi.uzh.ch/kitt/cores/>

Der Mann, der seinem Hund den Knochen wegnimmt, ist nachdenklich. Er legt ihn auf den Tisch. Dann geht er. (Dort liegt er)

Sentimentanalyse: <http://sentistrength.wlv.ac.uk/#Non-English>

Das Hotel war schlecht, die Zimmer waren hässlich, aber, immerhin, das Essen war gut.

“Der Minister kritisiert, dass der Präsident das Volk betrügt”

Gebiet:

Eine relativ neue Disziplin

- ▶ im Überschneidungsbereich von *Linguistik* und *Informatik*
- ▶ mit dem Ziel, menschliche *Sprache* mit dem Computer zu *verarbeiten*
- ▶ CL ist praxisrelevant
- ▶ CL ist Grundlagenforschung

Entwicklungsziele:

1. Entwicklung von *Anwendungsprogrammen*:

- ▶ Maschinelle Übersetzungssysteme
- ▶ Informationsextraktion, Sentimentanalyse, Text Mining
- ▶ Semantic Web und Web Mining
- ▶ Dialogsysteme (z.B. Flugreservierung)
- ▶ E-Learning-Applikationen (Sprachlernsoftware, Glossare, ...)
- ▶ Systeme zum Generieren geschriebener Texte
- ▶ Systeme zum Verstehen gesprochener Sprache
- ▶ Systeme zur Stilkontrolle, Grammar Checker, ...

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:

- ▶ Multilinguale (parallele) Baubanken und Korpuslinguistik
- ▶ Maschinelle Übersetzung
 - ▶ Automatische Übersetzung von Filmuntertiteln
 - ▶ Text und Berg: Kontextspezifische Übersetzung (Bilingwis)
- ▶ Relationen-Mining: Automatische Extraktion von Kernaussagen aus Texten
- ▶ Sentimentanalyse: Erkennen von Bewertungen, Meinungen, Stimmungen
- ▶ Sprachtechnologie für Barrierefreiheit
- ▶ **neu:** forensische Phonetik (Sprecheridentifikation bei der Verbrechensbekämpfung)

Computergestützte Textanalyse wird mittlerweile in vielen Disziplinen benötigt

- ▶ Psychologie:
Halbautomatische inhaltliche Analyse von transkribierten Therapiegesprächen
- ▶ Politikwissenschaften:
Halbautomatische Extraktion von Aussagen von Politiker/-innen über politische Akteure und Sachthemen
- ▶ Germanistik:
Bestimmen von Grundformen von flektierten Wörtern in einer digitalen Edition von Dokumenten aus dem Zürcher Sommer 1968
- ▶ Anglistik:
Syntaktische Analyse von riesigen Textsammlungen

Lehre in Zürich:

- ▶ Korpuslinguistik
- ▶ maschinelle Übersetzung
- ▶ Sentimentanalyse
- ▶ Einsatz von sprachtechnologischen Werkzeugen in der Linguistik
- ▶ computerlinguistische Annotation (z.B. Baumbanken)
- ▶ Vermittlung konkreter Programmierkenntnisse für die maschinelle Verarbeitung natürlicher Sprachen (keine Programmierkenntnisse vorausgesetzt)
- ▶ multilinguale Textanalyse
- ▶ **neu:** Erkennen gesprochener Sprache (forensische Phonetik)

Arbeitsbereiche/Branchen

- ▶ Universitäten
- ▶ Softwarehäuser, z.B. für maschinelle Sprachverarbeitung
- ▶ 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
- ▶ Spezialisierte Unternehmen, die grosse Datenmengen verarbeiten müssen, wie z.B. die pharmazeutische Industrie

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

Fach:

Seit WiSe 06/07 in Zürich (unter "Bologna") als

- ▶ Haupt- und Nebenfach
- ▶ je auf BA- und MA-Ebene (MA: seit HeSe 2009)

Zwei Studiengänge:

- ▶ Computerlinguistik (fällt 2020 weg)
- ▶ Computerlinguistik und Sprachtechnologie

Studienrichtung	Grösse	Umfang
Computerlinguistik (fällt 2020 weg)	NF	60
Computerlinguistik und Sprachtechnologie	NF	60
Computerlinguistik und Sprachtechnologie (fällt 2020 weg)	HF	90
ab 2020: Computerlinguistik und Sprachtechnologie	HF	120

Studierende der Philosophischen Fakultät

- ▶ BA als Hauptfach (90) bzw. Grosses (60)

Studierende anderer Fakultäten oder Hochschulen

- ▶ nach Regelung der betreffenden Institution, z.B.
 - ▶ als Nebenfach der Mathematisch-Naturwissenschaftlichen Fakultät

Anforderungen an einen CL-Studierenden

Computer

- ▶ keine besonderen

Programmieren

- ▶ Keine Vorkenntnisse erforderlich, aber Interesse!
- ▶ Python wird obligatorisch gelehrt
- ▶ Später können weitere Programmiersprachen gewählt werden

Fremdsprachen

- ▶ Bereitschaft, englischsprachige Fachtexte zu lesen
- ▶ kein Lateinobligatorium

Profil eines/r CL-Studierenden

Sprache

- ▶ Interesse an der automatischen Verarbeitung der menschlichen Sprache
- ▶ Spass am Basteln von sprachtechnologischen Anwendungen
- ▶ Spass am Umsetzen eigener Ideen zur Sprachverarbeitung

Gesellschaft

- ▶ Interesse an Fragen der modernen Informationsgesellschaft
- ▶ Mitwirken an der Flexibilisierung der Mensch-Maschine-Kommunikation

Informatik und Statistik

- ▶ Spass am Computer als mächtiges Modellierungsinstrument
- ▶ keine Scheu vor Statistik

Machinelles Lernen wird oft verwendet, ebenso statistische Verfahren man kann aber Anwender bleiben, muss nicht notgedrungen in die Theorie eintauchen

Die wichtigsten Informationen findet man auf unserer Homepage:

<http://www.cl.uzh.ch>

Fragen?

Zögern Sie nicht, stellen Sie Ihre Fragen ...