

## CL-Demos im World Wide Web

*Martin Volk, Universität Zürich  
(e-mail: volk@ifi.unizh.ch)*

Wie oft haben Sie schon nach einer Möglichkeit gesucht, mal eben schnell ein Morphologie-Programm oder einen Tagger für das Deutsche zu demonstrieren, wenn die Frage im Raum stand: Was ist denn eigentlich Computerlinguistik? Wie oft haben Sie schon bedauert, dass Sie ein Programm, von dem sie ausführlich gelesen hatten, nicht selbst ausprobieren können?

Durch die Anbindung von Programmen an das Internet können beide Probleme behoben werden. HTML-Formulare bieten die Möglichkeit, Wörter, Sätze oder ganze Texte in einem WWW-Browser einzugeben und über das Internet an ein verarbeitendes Programm weiterzuleiten. Viele Forscher machen auf diese Art mit online-Demo-Versionen auf ihre lauffähigen Programme aufmerksam, und Firmen bewerben so ihre Produkte.

Es ist jedoch bekanntlich schwierig, die gesuchte Stecknadel ‚NLP-Applikation‘ im Heuhaufen ‚Internet‘ zu finden. Und bei dieser speziellen Aufgabe helfen auch die Internet-Suchdienste nicht viel, denn interaktive NLP-Seiten finden sich unter sehr unterschiedlichen Bezeichnungen (wie z.B. ‚HPSG-Interaktiv‘ = ein Parser; ‚Interactive Demo of SVOX‘ = ein Sprachsynthesesystem; ‚NLP Parser Demonstration Page‘ = eine Auswahl unter mehreren Parsern und Parsingstrategien). Deshalb haben wir eine Sammlung mit **Interactive online CL demos** angelegt unter der Adresse: <http://www.ifi.unizh.ch/CL/InteractiveTools.html>.

Alle dort aufgeführten WWW-Links verweisen auf interaktive sprachverarbeitende Programme im Internet. Dabei bedeutet ‚interaktiv‘, dass der Benutzer eine Eingabe im WWW-Browser machen kann und diese durch das verarbeitende Programm unmittelbar analysiert und das Ergebnis zurückgeschickt wird. Unsere Sammlung beschränkt sich auf Systeme für Deutsch und Englisch und umfasst:

- Morphologie-Analyse von Wortformen
- Wortarten-Tagger zur kontext-abhängigen Zuteilung von Wortarten für einen oder mehrere Sätze
- Parser zur Syntaxanalyse einzelner Sätze
- Transformations-Systeme zur Umwandlung eines Aktiv-Aussage-Satzes in einen Passiv-Satz oder in einen Frage-Satz

- Generierung von Sätzen und kleinen Texten aus z. B. Börsentabellen oder Bildbeschreibungen
- Synthese gesprochener Sprache aus geschriebener Sprache
- Frage-Antwort-Systeme z. B. über geographische Fakten
- Maschinelle Übersetzung von einigen Sätzen oder ganzen WWW-Seiten

Programme, die lediglich das Nachschlagen in einem Lexikon ermöglichen, wurden nicht aufgenommen, jedoch enthält die Sammlung einige Systeme für online-Korporarecherchen (unter Auflösung von Flexion und anderen Wortbildungen) und für Recherchen in semantischen Netzwerken. Verweise auf Grammatik- oder Stilprüfprogramme fehlen, da dazu bisher keine interaktiven WWW-Angebote gefunden werden konnten.

Schließlich gibt es noch einen Abschnitt mit „General Tools“. Dort finden sich z. B. Sprachidentifizierungsprogramme, also Programme, die aufgrund eines oder mehrerer eingegebener Sätze entscheiden können, um welche Sprache es sich handelt.

Wir hoffen, dass die Aufstellung für viele von Ihnen nützlich ist. Um die Sammlung aktuell halten zu können, sind wir darauf angewiesen, dass Sie uns Hinweise auf neue interaktive CL-Software (und natürlich auch auf alte ungültige Links) schicken. Bitte senden Sie Ihre Vorschläge an Martin Volk, Universität Zürich (volk@ifi.unizh.ch).