

Kursplan för höstterminen 2003

SPRÅKBEHANDLING OCH DATALINGVISTIK

EDA171

Language Processing and Computational Linguistics

Antal poäng: 5. **Betygskala:** UG. **Valfri för:** D4, E4, F4.
Kursansvarig: Professor Pierre Nugues, Pierre.Nugues@cs.lth.se,
Inst f datavetenskap. **Förkunskapskrav:** EDA027 Algoritmer och
datastrukturer. **Prestationsbedömning:** Godkända
inlämningsuppgifter, projekt och godkända rapporter. **Hemsida:**
<http://www.cs.lth.se>. **Övrigt:** Kursen kan komma att ges på
engelska. Begränsat deltagarantal, totalt 40 platser (LTH och
LU). Kursen ges gemensamt för studenter vid LTH, vid
naturvetenskaplig fakultet och för studenter/forskarstuderande i
Lingvistik Om det blir fler sökande än platser fördelas platserna
mellan grupperna proportionellt mot antalet sökande.
Urvalsgrunder för platser som tillfaller studenter vid LTH blir: 1.
Antal poäng som uppnåtts vid LTH. 2. Antal poäng i kurser inom
Datavetenskap.

Mål

Kursen introducerar teorier och tekniker som tillämpas i maskinell
behandling av naturligt språk (Natural Language Processing, NLP).
Den täcker hela fältet, från taligenkänning och talsyntes till
semantik och dialogsystem.

Kursen inriktar sig på tillämpningar i industriell skala eller på
försöksnivå, såsom documentsökning på Internet,
informationsextrahering, talande agentprogram och verbal
interaktion i virtuella världar. Centrala algoritmer beskrivs i Prolog
eller med reguljära uttryck.

Innehåll

- En översikt över NLP: Presentation av NLP, tillämpningar,
lingvistikens delområden, exempel
- Korpuslingvistik: Reguljära uttryck, ändliga automater,
introduktion till Perl, konkordanser, tokenisering,
frekvenslistor, kollokationer
- Morfologi och ordklasstagning: Morfologi, transduktorer,
ordklasstagning
- Frässtrukturgrammatiker i Prolog: Konstituent, syntaxträd,
prolog för analys av naturligt språk, DCG-regler, variabler,

syntaktisk parsning, semantisk parsning med kompositionell analys

- Syntaktiska formalismer: Konstituens och dependens, chart parsing, statistisk parsning, funktioner, dependensparsning
- Semantik: Formell semantik, lambda-kalkyl, kompositionalitet; substantiv, verb, determinerare, ord och betydelse, lexikal semantik, kasusgrammatiker, semantiska grammatikor
- Diskurs och dialog: diskurs och retorik, anaforer, dialogstruktur, RST (rhetorical structure theory), dialog, ändliga automater, adjacency pairs, talhandlingar, multimodalitet
- Översikt över talsyntes och taligenkänning

Litteratur

Kompendium, som säljs av institutionen.

Rekommenderad referenslitteratur:

Manning and Schütze: Foundations of Statistical Natural Language Processing, MIT Press, 1999.

Jurafsky, Martin, Vander Linden, Kehler: Speech and Language Processing, Prentice Hall, 2000.