



Kursplan för läsåret 2010/2011
(Genererad 2010-06-28.)

SPRÅKTEKNOLOGI

Language Technology

EDAN20

Antal högskolepoäng: 7,5. **Betygsskala:** TH. **Nivå:** A (Avancerad nivå). **Huvudområde:** Teknik. **Undervisningsspråk:** Kursen kan komma att ges på engelska. **Överlappar följande kurs/kurser:** EDA171. **Valfri för:** C4, D4, D4pv. **Kursansvarig:** Professor Pierre Nugues, Pierre.Nugues@cs.lth.se, Inst f datavetenskap. **Förkunskapskrav:** EDAA01 Programmeringsteknik - fördjupningskurs eller EDA027 Algoritmer och datastrukturer. **Prestationsbedömning:** För betyg 3 krävs godkända inlämningsuppgifter (laborationer). För möjlighet till högre betyg krävs även skriftlig tentamen. Inlämningsuppgifterna (laborationerna) skall utföras i grupper om två eller tre studenter, men kan även göras individuellt. Den första laborationen ägnas åt de programmeringsverktyg som används i kursen. De fem övriga uppgifterna behandlas under de följande laborationstillfällena. **Hemsida:** <http://cs.lth.se/>.

Syfte

Under de senaste 15 åren har de språkteknologiska metoderna mognat avsevärt på grund av att mängden tillgänglig skriftlig och talad information har ökat kraftigt, vilket har lett till ökande behov av att behandla den automatiskt. Fastän de flesta datorsystem inte enbart ägnar sig åt språkbehandling så finns det numera många applikationer som i någon mån är "språkmedvetna" och har inbyggda språkteknologiska funktioner såsom stavnings- och grammatikkontroll, sökning och sammanfattning av information eller ett talbaserat dialoggränssnitt. Detta gör att fältet är av ökande betydelse för CS-ingenjörer.

Kursen ger en inledning till språkteknologins metoder. Den försöker täcka hela fältet från teckenkodning och statistiska språkmodeller till syntaktisk analys, semantik och dialogsystem. Kursen inriktar sig på välbeprövade tillämpningar i industriell skala eller på försöksnivå.

Mål

Kunskap och förståelse

För godkänd kurs skall studenten

- Förstå fältet av språkteknologiska metoder och viktiga applikationer som använder dem
- Känna till de viktigaste teknikerna, grundläggande algoritmer och allmänna arkitekturer som används i applikationer
- Skapa och konstruera språkbehandlingsalgoritmer. Skriva, tolka, utvärdera och förbättra

dem under laborationerna.

Färdighet och förmåga

För godkänd kurs skall studenten

- Förstå och utveckla annoteringsscheman, skapa och bearbeta strukturerade dokument genom att använda XML
- Förstå och skriva reguljära uttryck och använda dem i programmeringsspråk som Perl eller Java
- Använda logik och logikprogrammeringsspråk som Prolog- eller beskrivningslogik
- Förstå och använda maskininlärningsalgoritmer och statistiska tekniker
- Utveckla och utvärdera algoritmer i de viktiga fälten av språkteknologier, språkmodellering, partiell parsning, dependensparsning, genom att använda verkliga data

Värderingsförmåga och förhållningssätt

För godkänd kurs skall studenten

- Visa nyfikenhet, kreativitet och förmåga till problemlösning
- Visa förståelse för industriella och forskningsrelaterade frågor i språkteknologiområdet

Innehåll

- *En översikt över språkteknologi*: delområden, tillämpningar och exempel.
- *Korpuslingvistik*: Reguljära uttryck, ändliga automater, introduktion till Perl, konkordanser, tokenisering, frekvenslistor, kollokationer.
- *Morfologi och ordklasstagning*: Morfologi, transduktorer, ordklasstagning.
- *Frasstrukturgrammatiker i Prolog*: Konstituent, syntaxträd, DCG-regler, unifiering.
- *Partiell parsning*: Detektering av multiord, detektering av substantivgrupp och verbgrupp, informationsextrahering, utvärdering.
- *Syntaktiska formalismer*: Konstituens och dependens, funktioner. Parsning, statistisk parsning, dependensparsning.
- *Semantik*: Formell semantik, lambda-kalkyl, lexikal semantik, predikat-argumentstruktur, ramsemantik, semantisk parsning.
- *Diskurs och dialog*: Referens och koreferens, diskurs och retorik, diskursrelationer, analys av diskursrelationer, dialog, talhandling, multimodalitet.

Litteratur

Nugues Pierre, An Introduction to Language Processing with Perl and Prolog. An Outline of Theories, Implementation, and Application with Special Consideration of English, French, and German. Series: Cognitive Technologies, Springer Verlag, 2006, ISBN: 3-540-25031-X.