

DATN06, Datavetenskap: Språkbehandling och datalingvistik, 7,5 högskolepoäng

*Computer Science: Language Processing and Computational
Linguistics, 7.5 credits*

Avancerad nivå / Second Cycle

Fastställande

Kursplanen är fastställd av Naturvetenskapliga fakultetens utbildningsnämnd 2007-01-31 att gälla från och med 2007-07-01, höstterminen 2007.

Allmänna uppgifter

Kursen är en valbar kurs på avancerad nivå för en naturvetenskaplig masterexamen med huvudområdet datavetenskap. Kursen ges även som fristående kurs.

Undervisningsspråk: Engelska och Svenska
Kursen ges eventuellt på engelska.

Huvudområde

Datavetenskap

Fördjupning

A1N, Avancerad nivå, har endast kurs/er på grundnivå som förkunskapskrav

Kursens mål

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.

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.

- Utforma, utveckla, utvärdera och beskriva en fullständig språkteknologisk prototyp.

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, semantiska nät och dialog, genom att använda verkliga data.
 - Konstruera en färdig språkteknologisk prototyp i ett grupprojeckt med två eller tre deltagare (eventuellt en).
- Söka litteratur och elektroniskt publicerat material inom språkteknologiområdet.
 –Utvärdera och reflektera över existerande system och forskning inom området.
 –Skapa och implementera prototypen.
 –Skriftligt och muntligt presentera projektet.

Värderingsförmåga och förhållningssätt

För godkänd kurs skall studenten:

- Visa nyfikenhet, kreativitet och förmåga till problemlösning.
- Visa förståelse för industriella och forsknings relaterade frågor i språkteknologiområdet.

Kursens 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, DCG-regler, variabler, syntaktisk parsning, semantisk parsning med kompositionell analys.
- Partiell parsning och informationsextrahering.

- Syntaktiska formalismer: Konstituens och dependens, funktioner.
- Parsning: Chart parsing, statistisk parsning, 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 röstigenkänning.

Kursens genomförande

Undervisningen utgörs av föreläsningar och gruppövningar.

Kursens examination

Examination sker med inlämningsuppgifter, projekt och uppsatser. Inlämningsuppgifter och projekt utförs normalt i grupper om två eller tre studenter, men kan även utföras individuellt. Projektet handleds av en lärare och redovisas i slutet av kursen. Om projektet inte godkänns vid detta tillfälle ges en möjlighet att inom en månad förbättra resultatet och presentera på nytt.

Provlmoment för denna kurs finns i en bilaga i slutet av dokumentet.

Betyg

Betygsskalan omfattar betygsgraderna Underkänd, Godkänd, Väl godkänd. Betygsgraderna på kursen är väl godkänd, godkänd och underkänd. För godkänt betyg på kursen krävs godkända inlämningsuppgifter, godkänt projekt och godkänd uppsats.

Förkunskapskrav

För tillträde till kursen krävs kunskaper motsvarande DAT501, Datalogi, allmän kurs (1-20), 20 poäng.

Övrigt

Kursen kan inte tillgodoräknas i examen tillsammans med DAT171, Språkbehandling och datalingvistik, 5 poäng.

Prov/moment för kursen DATN06, Datavetenskap: Språkbehandling och datalingvistik

Gäller från H07

0701 Språkbehandling och datalingvistik, 7,5 hp
Betygsskala: Underkänd, Godkänd, Väl godkänd