



Kursplan

för kurs på avancerad nivå

Språkteknologi

Language Engineering

6.0 Höskolepoäng

6.0 ECTS credits

Kurskod: DA7032
Gäller från: HT 2011
Fastställd: 2011-01-18
Institution Matematiska institutionen

Huvudområde: Datalogi
Fördjupning: A1N - Avancerad nivå, har endast kurs/er på grundnivå som förkunskapskrav

Beslut

Denna kursplan är fastställd av Naturvetenskapliga fakultetsnämnden vid Stockholms universitet 2011-01-18.

Beslut om upphävande är fattat av Områdesnämnden för naturvetenskap 2019-10-07.

Förkunskapskrav och andra villkor för tillträde till kursen

För tillträde till kursen krävs kunskaper motsvarande kurserna Objektorienterad programmering, GN, 7,5 hp (DA3002) och Sannolikhetsteori I, GN, 7,5 hp (MT3001) eller 60 hp inom data- och systemvetenskap. Svenska B.

Kursens uppläggning

Provkod	Benämning	Höskolepoäng
THEO	Teori	3
ILUP	Inlämningsuppgift	1.5
LABO	Laborationer	1.5

Kursens innehåll

a. Kursen behandlar: Olika metoder för analys, generering och filtrering av mänskligt språk med fokus på text.

Teori: Språkteknologins historiska utveckling och grunder, morfologi, syntax och semantik, vektorrumsmodeller, utvärderingsmetoder, terminologilära, maskininlärning, informationsteori och Markovmodeller, algoritmer och datastrukturer för kompakt lagring och uppslagning i lexikon.

Tekniker: Morfologisk analys och generering, språkstatistik och korpusbearbetning, parsning, språkgenerering, ordklasstagning, namnigenkänning och probabilistisk parsning, statistisk lexikal semantik.

Tillämpningsområden: Stavnings- och grammatikkontroll, informationssökning, ordprediktion för smart textinmatning, textklustrering och textkategorisering, datorstött språkinlärning, dialogsystem, textsammanfattning, talteknologi, lokalisering och internationalisering.

b. Kursen består av följande moment:

- Teori (Theory), 3 hp
- Laborationer (Practical Exercises), 1,5 hp
- Inlämningsuppgift (Home Assignment), 1,5 hp

Förväntade studieresultat

Efter att ha genomgått kursen ska studenten kunna:

- förklara och använda begrepp inom språkvetenskapens grundläggande nivåer såsom morfologi, syntax, semantik, diskurs och pragmatik
- tillämpa kunskap om morfologi, syntax och lexikal semantik för att bygga språkteknologiska system samt förklara uppbyggnaden av existerande system utifrån dessa nivåer
- klargöra skillnaderna mellan analys, generering och filtrering med avseende på textbaserade system
- använda grundläggande verktyg inom språkteknologin såsom ordklassstaggare, frasanalysverktyg samt olika typer av korpusar och lexikon för att kunna bygga egna program
- förklara och använda standardmetoder inom språkteknologin som bygger på såväl regler som statistik och maskininlärning
- praktiskt tillämpa metoder som bygger på ändliga automater/transduktorer, kontextfri grammatik, ordfrekvenser, n-gram samförekomststatistik, Markov-modeller och vektorrumsmodeller
- analysera och förklara vilka språkteknologiska problem som går att lösa med tillfredställande resultat samt vilka som ligger bortom forskningshorisonten
- förklara i detalj hur en stavningskontroll, grammatikkontroll, någon typ av taggning med hjälp av maskininlärning, stemmer och en algoritm för statistisk utvinning av relaterade ord fungerar
- utforma och genomföra enklare utvärderingar av något språkteknologiskt system samt tolka resultaten
- självständigt lösa ett välavgränsat praktiskt språkteknologiskt problem eller analysera det genom teoretiska studier

Undervisning

Undervisningen består av föreläsningar och laborationer.

Deltagande i laborationer och därmed integrerad gruppundervisning är obligatoriskt. Om särskilda skäl föreligger kan examinator efter samråd med vederbörande lärare medge den studerande befrielse från skyldigheten att delta i vissa obligatoriska moment.

Kunskapskontroll och examination

a. Kursen examineras på följande vis: Kunskapskontroll av momentet Teori sker genom skriftligt prov och av momentet Inlämningsuppgift genom skriftlig och muntlig redovisning.

b. Betygssättning sker enligt sjugradig målrelaterad betygsskala:

A = Utmärkt

B = Mycket bra

C = Bra

D = Tillfredsställande

E = Tillräckligt

Fx = Otillräckligt

F = Helt otillräckligt

c. Kursens betygsriterier delas ut vid kursstart.

d. För godkänt krävs lägst betygsgraden E samt godkänt på momentet Laborationer och deltagande i all obligatorisk undervisning.

e. Studerande som underkänts i ordinarie prov har rätt att genomgå minst fyra ytterligare prov så länge kursen ges. Med prov jämföras också andra obligatoriska kursdelar. Studerande som godkänts på prov får inte genomgå förnyat prov för högre betyg. Studerande som underkänts på prov två gånger har rätt att begära att annan examinator utses vid nästkommande prov. Framställan härom ska göras till institutionsstyrelsen. Institutionen tillämpar inte någon begränsning av den tid under vilken omexamination står öppen för studenten.

f. Vid betyget Fx kan examinator besluta att möjlighet till komplettering upp till betyget E ska ges. Besked om möjligheten kan ges lämnas i kursens betygsriterier. Examinator beslutar om vilka kompletteringsuppgifter som ska utföras. Kompletteringen ska äga rum före nästa examinationstillfälle.

Övergångsbestämmelser

Studerande kan begära att examination genomförs enligt denna kursplan även efter det att den upphört att gälla, dock högst tre gånger under en tvåårsperiod efter det att undervisning på kursen upphört. Framställan härom ska göras till institutionsstyrelsen.

Begränsningar

Kursen kan ej ingå i examen tillsammans med kursen Språkteknologi, AN, 6 hp (DA7022) eller Språkteknologi för datorlingvister, GN, 7,5 hp (DA3010), eller motsvarande.

Övrigt

Kursen ges som fristående kurs, men kan även ingå i masterprogrammet i datalogi.

Kurslitteratur

Kurslitteratur beslutas av institutionsstyrelsen och redovisas därefter i bilaga till kursplanen.