Matematiska institutionen

Stockholms universitet

Kursplan

för kurs på grundnivå

Språkteknologi för lingvister **7.5 Högskolepoäng**
Language Engineering for Linguists **7.5 ECTS credits**

Kurskod:	DA3017
Gäller från:	HT 2015
Fastställd:	2014-08-22
Institution	Matematiska institutionen
Huvudområde:	Datalogi
Fördjupning:	G1F - Grundnivå, har mindre än 60 hp kurs/er på grundnivå som förkunskapskrav

Beslut

Denna kursplan är fastställd av Naturvetenskapliga fakultetsnämnden vid Stockholms universitet 2014-08-22.
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 Lingvistik IV 30 hp (LIN400) samt Programmeringsteknik för lingvister, 7,5 hp (DA3016).

Kursens uppläggning

Provkod	Benämning	Högskolepoäng
THEO	Teori	3.5
LABO	Laborationer	2
ILUP	Inlämningsuppgift	2

Kursens innehåll

a. Kursen behandlar: Olika metoder för analys, generering och filtrering av mänskligt språk med fokus på text. Regelbaserade och statistiska metoder används och undersöks till exempel för informationssökning, stavnings- och grammatikkontroll, och textsammanfattning.
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, ordklasstaggning, namnigenkänning och probabilistisk parsning, statistisk lexikal semantik.
Tillämpningsområden: Stavnings- och grammatikkontroll, informationssökning, ordprediktion för smart textinmatning, textklustring och textkategorisering, datorstödd språkinlärning, dialogsystem, textsammanfattning, talteknologi, lokalisering och internationalisering.

b. Kursen består av följande moment:
•Teori (Theory), 3,5 hp
•Laborationer (Practical Exercises), 2 hp
•Inlämningsuppgift (Home Assignment), 2 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 ordklasstaggare, 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, Markovmodeller 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 viss obligatorisk undervisning.

Kunskapskontroll och examination

a. Kursen examineras på följande vis: Kunskapskontroll av momentet Teori sker genom skriftligt och prov och av momenten Inlämningsuppgift respektive Laborationer 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

Betygssättning av moment Laborationer 2 hp sker enligt tvågradig betygsskala: godkänd (G) eller underkänd (U).

c. Kursens betygskriterier 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 ytterligare prov så länge kursen ges. Antalet provtillfällen är inte begränsat. Med prov jämställs 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 lärare utses för att bestämma betyg på kursen. Framställan härom ska göras till institutionsstyrelsen.
Kursen har minst två examinationstillfällen för varje moment per läsår de år då undervisning ges. Mellanliggande år ges minst ett examinationstillfälle.

f. Vid betyget Fx ges möjlighet att komplettera upp till betyget E. Examinator beslutar om vilka kompletteringsuppgifter som ska utföras och vilka kriterier som ska gälla för att bli godkänd på kompletteringen. 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. Bestämmelsen gäller även vid revidering av kursplanen.

Begränsningar

Kursen kan ej ingå i examen tillsammans med kurserna Språkteknologi (DA7032), Språkteknologi för datorlingvister (DA3011), eller motsvarande.

Övrigt

Kursen ingår i kandidatprogrammet i lingvistik.

Kurslitteratur

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