



DH2418 Språkteknologi 6,0 hp

Language Engineering

Fastställande

Kursplan för DH2418 gäller från och med HT09

Betygsskala

A, B, C, D, E, FX, F

Utbildningsnivå

Avancerad nivå

Huvudområden

Datalogi och datateknik, Informationsteknik, Informations- och kommunikationsteknik

Särskild behörighet

För fristående kursstuderande krävs 90 högskolepoäng varav 45 högskolepoäng inom matematik eller informationsteknik. Dessutom krävs svenska B eller motsvarande och engelska A eller motsvarande.

Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

Studenterna skall efter kursen 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 ordklasstaggar, 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.

För att kunna:

- arbeta på språkteknologiska företag.
- fortsätta med språkteknologiskt orienterade studier.
- göra ett examensarbete inom datalogi eller människa-datorinteraktion med en språkteknologisk inriktning.
- vara en viktig länk mellan systemarkitekter, programmerare och interaktionsdesigner i såväl arbetsliv som forskning.

Kursinnehåll

Teori:

Språkteknologins historiska utveckling och grunder, morfologi, syntax och semantik, vektorrumsmodeller, utvärderingsmetoder, terminologilära, maskininlärning, informationsteori och Markov-modeller, 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, textklustrering och textkategorisering, datorstödd språkinlärning, dialogsystem, textsammanfattning, talteknologi, lokalisering och internationalisering.

Kurslitteratur

Kurslitteratur meddelas på kursens hemsida senast 4 veckor före kursstart.

Föregående kursomgång användes Jurafsky & Martin, Speech and language processing samt material producerat vid institutionen.

Examination

- INLA - Inlämningsuppgift, 1,5 hp, betygsskala: A, B, C, D, E, FX, F
- TEN2 - Tentamen, 3,0 hp, betygsskala: A, B, C, D, E, FX, F
- LAB2 - Laborationsuppgifter, 1,5 hp, betygsskala: P, F

Examinator beslutar, baserat på rekommendation från KTH:s handläggare av stöd till studenter med funktionsnedsättning, om eventuell anpassad examination för studenter med dokumenterad, varaktig funktionsnedsättning.

Examinator får medge annan examinationsform vid omexamination av enskilda studenter.

När kurs inte längre ges har student möjlighet att examineras under ytterligare två läsår.

I denna kurs tillämpas skolans hederskodex, se: <http://www.kth.se/csc/student/hederskodex>.

Övriga krav för slutbetyg

En tentamen (TEN2; 3 hp)

Laborationsuppgifter (LAB2; 1,5 hp)

Inlämningsuppgift (INLA; 1,5 hp).

Etiskt förhållningssätt

- Vid grupparbete har alla i gruppen ansvar för gruppens arbete.
- Vid examination ska varje student ärligt redovisa hjälp som erhållits och källor som använts.
- Vid muntlig examination ska varje student kunna redogöra för hela uppgiften och hela lösningen.