



DD2418 Språkteknologi 6,0 hp

Language Engineering

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

Fastställande

Kursplan för DD2418 gäller från och med HT12

Betygsskala

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

Utbildningsnivå

Avancerad nivå

Huvudområden

Datalogi och datateknik, Informations- och kommunikationsteknik, Informationsteknik

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 språk teknologiska 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 språk teknologin såsom ordklasstaggare, frasanalysverktyg samt olika typer av korpusar och lexikon för att kunna bygga egna program,
- förklara och använda standardmetoder inom språk teknologin 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 språk teknologiska 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 språk teknologiskt system samt tolka resultaten,
- självständigt lösa ett välavgränsat praktiskt språk teknologiskt problem eller analysera det genom teoretiska studier,

för att kunna:

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

Kursinnehåll

Teori:

Språk teknologins 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, textklustring 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

- INL1 - Inlämningsuppgift, 1,5 hp, betygsskala: A, B, C, D, E, FX, F
- LAB1 - Laborationer, 1,5 hp, betygsskala: P, F
- TEN1 - Tentamen, 3,0 hp, betygsskala: A, B, C, D, E, FX, 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 enstaka studenter.

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

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.