



DD1418 Språkteknologi med in- troduktion till maskininlärning 6,0 hp

Language Engineering with Introduction to Machine Learning

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

Fastställande

Kursplan för DD1418 gäller från och med VT19

Betygsskala

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

Utbildningsnivå

Grundnivå

Huvudområden

Teknik

Särskild behörighet

En kurs i datalogi motsvarande DD1320/DD1321 Tillämpad datalogi, DD1327 Grundläggande datalogi eller DD1338 Algoritmer och datastrukturer. En kurs i sannolikhets teori och statistik motsvarande SF1901.

Undervisningspråk

Undervisningspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

Efter godkänd kurs ska studenterna 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,
- använda grundläggande verktyg inom språkteknologin såsom ordklasstagare, 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 tillfredsställande resultat samt vilka som ligger bortom forskningshorisonten,
- förklara 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,

i syfte att:

- göra ett kandidatexamensarbete med språkteknologisk eller maskininlärningsinriktning,
- 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, maskininlärning, informationsteori och Markov-modeller.

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, textklustring och textkategorisering, datorstödd språkinlärning, dialogsystem, talteknologi och maskinöversättning.

Kursupplägg

Teoriföreläsningar och tillämpningsföreläsningar varvade med praktiska laborationer. Ett avslutande projektarbete som redovisas muntligt och skriftligt.

Kurslitteratur

Meddelas senast 10 veckor före kursstart på kurswebben.

Examination

- LAB1 - Laborationsuppgifter, 1,5 hp, betygsskala: P, F
- PRO1 - Projektuppgift, 1,5 hp, betygsskala: A, B, C, D, E, FX, 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.

Om särskilda skäl finns kan annan examinationsform användas.

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

Övriga krav för slutbetyg

Godkänd labbkurs, projektuppgift och tenta.

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.