



INSTITUTIONEN FÖR FILOSOFI, LINGVISTIK OCH VETENSKAPSTEORI

LT2124 Teman i datalingvistik och språkteknologi, 7,5 högskolepoäng
Themes in NLP and language technology, 7.5 credits
Avancerad nivå / Second Cycle

Fastställande

Kursplanen är fastställd av Institutionen för filosofi, lingvistik och vetenskapsteori 2020-11-23 att gälla från och med 2021-01-18, vårterminen 2021.

Utbildningsområde: Naturvetenskapligt 100 %

Ansvarig institution: Institutionen för filosofi, lingvistik och vetenskapsteori

Inplacering

Kan även ges som fristående kurs.

Kursen kan ingå i följande program: 1) Språkteknologi, masterprogram (H2MLT)

Huvudområde

Språkteknologi

Fördjupning

A1N, Avancerad nivå, har endast kurs/er på grundnivå som förkunskapskrav

Förkunskapskrav

För tillträde till kursen krävs:

Motsvarande kandiexamen i

- datavetenskap, datalingvistik eller språkteknologi,
- lingvistik (inkl minst 30 hp formell lingvistik),
- angränsande ämne, tex kognitionsvetenskap, språk, filosofi eller matematik (i dessa fall krävs motsvarande 30 hp inom antingen formell lingvistik eller programmering), eller
- motsvarande efter bedömning.

Godkända kunskaper i engelska motsvarande gymnasiekursen Engelska 5/Engelska A.

Lärandemål

Kunskap och förståelse

- redogöra för för- och nackdelar med olika angreppssätt inom datalingvistik,
- redogöra för vanliga parsningsalgoritmer,
- redogöra för teorierna bakom olika datadrivna angreppssätt,
- förklara skillnaden mellan metoder och tillämpningar inom datalingvistik,

Färdigheter och förmåga

- tillämpa samtida metoder för grundläggande datalingvistiska problem,
- använda olika mjukvaruverktyg för att lösa datalingvistiska problem med hjälp av annoterade korpora,
- skriva enkla program som tillämpar datalingvistiska verktyg och mjukvarubibliotek,

Värderingsförmåga och förhållningssätt

- välja mellan olika tekniker och olika representationer av data för att lösa specifika datalingvistiska problem,
- välja mellan existerande tillämpningar och tekniker i syfte att tillämpa dem på ett nytt problem,
- välja lämpliga egenskaper för att utvärdera en föreslagen lösning av ett NLP-problem.

Innehåll

Kursen ger en överblick över tillämpningar och tekniker inom datalingvistik (eller 'Natural Language Processing', NLP), de huvudsakliga problemen inom datalingvistik, och några standardlösningar på dessa problem.

Parallellt får studenterna en praktisk programmeringserfarenhet av att lösa dessa uppgifter. Vi använder samma programmeringsspråk som på LT2001 (Introduktion till programmering) tillsammans med standardbibliotek för NLP samt kommandoradsverktyg.

Kursen är uppdelad i tre block; ett tar upp grundläggande begrepp och de andra täcker delområden inom NLP.

1. Grundläggande begrepp:

- övervakad inlärning (eng. 'supervised learning')
- utvärdering
- symboliska och statistiska metoder

2. Ord och meningar:

- korpora och korpusannotering
- finite-state-metoder för segmentering och morfologisk analys
- statistisk språkmodellering med n-gram Markovmodeller
- vektorrumsrepresentationer och relaterade operationer

3. Översikt över vanliga NLP-problem, inkluderande ett urval av följande:

- ordklasstagning
- ordbetydelsedisambiguering
- maskinöversättning
- distributionell semantik
- textklassificering
- textsättning av bilder

Former för undervisning

Föreläsningar, förevisningar, och handledda labtillfällen för hjälp med uppgifter.

Undervisningsspråk: engelska

Former för bedömning

Kursen examineras genom programmeringsorienterade inlämningsuppgifter, skrivna inlämningsuppgifter, och skriftliga och/eller webbaserade prov. Obligatorisk närvaro kan gälla för vissa kursmoment.

Student äger rätt till byte av examinator efter att ha underkänts två gånger på samma examination, om det är praktiskt möjligt. En sådan begäran ställs till institutionen och skall vara skriftlig. Antal provtillfällen är begränsade till fem stycken.

Rättande lärare kan begära komplettering av examination.

Betyg

På kursen ges något av betygen Väl godkänd (VG), Godkänd (G) och Underkänd (U). För betyget Godkänd krävs:

- Godkända inlämningsuppgifter
- Godkända skrivna eller webbaserade test, om sådana utdelas, inklusive eventuellt skriftligt prov

För betyget Väl godkänd krävs:

- Väl godkända prestationer i inlämningsuppgifter

- Godkända skrivna eller webbaserade test, om sådana utdelas, inklusive eventuellt skriftligt prov

Kursvärdering

Studenter som deltar i eller har avslutat kursen ges möjlighet att anonymt framföra erfarenheter av och synpunkter på denna i en kursvärdering. En sammanställning av kursvärderingen och kursansvarig lärarens reflektion tillgängliggörs för studenterna inom rimlig tid efter kursslut. Nästa gång kursen ges presenteras sammanställningen och eventuella genomförda åtgärder.