



INSTITUTIONEN FÖR FILOSOFI, LINGVISTIK OCH VETENSKAPSTEORI

LT2003 Grundläggande språkteknologi, 15 högskolepoäng

Natural language processing, 15 credits

Avancerad nivå / Second Cycle

Fastställande

Kursplanen är fastställd av Institutionen för filosofi, lingvistik och vetenskapsteori 2018-05-29 och senast reviderad 2019-05-31. Den reviderade kursplanen gäller från och med 2019-09-02, höstterminen 2019.

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 kandidatexamen i något av följande:

- datavetenskap,
 - datalingvistik eller språkteknologi,
 - lingvistik (inklusive motsvarande 30 hp formell lingvistik eller programmering),
 - angränsande ämne, tex kognitionsvetenskap, språk, filosofi eller matematik, inklusive motsvarande 30 hp formell lingvistik eller programmering
- eller motsvarande efter bedömning.

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

Lärandemål

Kunskap och förståelse

- redogöra för grundläggande begrepp inom NLP, automatteori, sannolikhetsteori, linjär algebra, och matematik
- redogöra för för- och nackdelar med olika angreppssätt, t.ex. övervakad/oövervakad inlärning, olika algoritmer, etc,
- redogöra för olika utvärderingsmetoder och deras användningsområden,
- redogöra för vanliga parsningsalgoritmer,
- redogöra för teorierna bakom olika datadrivna angreppssätt,
- veta vad plagiering innebär,
- ha kunskap om GU:s regler om plagiering,

Färdigheter och förmåga

- tillämpa samtida metoder för grundläggande NLP-problem som t ex ordklasstagning,
- använda olika maskininlärningstekniker för att lösa NLP-problem med hjälp av annoterade korpora,
- använda grundläggande maskininlärning och relaterade programmeringsmetoder inom NLP,
- använda kommandoradsverktyg för att hantera data,
- använda statistiska utvärderingsmetoder,
- identifiera plagiering,
- hitta information om plagiering och GU:s regler om plagiering,

Värderingsförmåga och förhållningssätt

- utvärdera och välja mellan olika maskininlärningstekniker och olika representationer av data för att lösa specifika NLP-problem,
- välja lämpliga egenskaper för att utvärdera en föreslagen lösning av ett NLP-problem.

Innehåll

Kursen ger en teoretisk ingång till de huvudsakliga problemen inom 'Natural Language Processing', och några av deras standardlösningar.

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 och

kommandoradsverktyg.

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

1. Grundläggande begrepp:

- grundbegrepp inom NLP
- grundläggande automatteori och matematisk lingvistik
- grundläggande sannolikhets teori och maskininlärning
- algebra och annan relevant matematik
- utvärderingsmått, inkluderande korrekthet, precision och täckning

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
- dialogsystem
- distributionell semantik
- textklassificering

Former för undervisning

Föreläsningar, förevisningar, och handledda tillfällen 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.

Komplettering av examinerad studentprestation medges.

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 lärares reflektion tillgängliggörs för studenterna inom rimlig tid efter kursslut. Nästa gång kursen ges presenteras sammanställningen och eventuella genomförda åtgärder.