



LUNDS UNIVERSITET
Lunds Tekniska Högskola

Kursplan för

Språkteknologi

Language Technology

EDAN20, 7,5 högskolepoäng, A (Avancerad nivå)

Gäller för: Läsåret 2019/20

Beslutad av: Programledning C/D

Beslutsdatum: 2019-04-01

Allmänna uppgifter

Valfri för: C4-pv, D4-pv, D4-mai, E4-bg, F4, F4-pv, F4-mai, Pi4-pv, Pi4-bam

Undervisningsspråk: Kursen ges på engelska

Syfte

Under de senaste 15 åren har de språkteknologiska metoderna mognat avsevärt på grund av att mängden tillgänglig skriven och talad information har ökat kraftigt, vilket har lett till ökande behov av att behandla den automatiskt. Fastän de flesta datorsystem inte enbart ägnar sig åt språkbehandling så finns det numera många applikationer som i någon mån är "språkmedvetna" och har inbyggda språkteknologiska funktioner såsom stavnings- och grammatikkontroll, sökning och sammanfattning av information eller ett talbaserat dialoggränssnitt. Detta gör att fältet är av ökande betydelse för CS-ingenjörer.

Kursen ger en inledning till språkteknologins metoder. Den försöker täcka hela fältet från teckenkodning och statistiska språkmodeller till syntaktisk analys, semantik och dialogsysteem. Kursen inriktar sig på välbeprövade tillämpningar i industriell skala eller på försöksnivå.

Mål

Kunskap och förståelse

För godkänd kurs skall studenten

- Förstå fältet av språkteknologiska metoder och viktiga applikationer som använder dem
- Känna till de viktigaste teknikerna, grundläggande algoritmer och allmänna arkitekturer som används i applikationer
- Skapa och konstruera språkbehandlingsalgoritmer. Skriva, tolka, utvärdera och förbättra dem under laborationerna.

Färdighet och förmåga

För godkänd kurs skall studenten

- Förstå och utveckla annoteringsscheman, skapa och bearbeta strukturerade dokument
- Förstå och skriva reguljära uttryck och använda dem i programmeringsspråk som Python
- Förstå och använda maskininlärningsalgoritmer och statistiska tekniker
- Utveckla och utvärdera algoritmer i de viktiga fälten av språkteknologier, språkmodellering, partiell parsning, dependensparsning, semantisk parsning genom att använda verkliga data

Värderingsförmåga och förhållningssätt

För godkänd kurs skall studenten

- Visa nyfikenhet, kreativitet och förmåga till problemlösning
- Visa förståelse för industriella och forskningsrelaterade frågor i språkteknologiområdet

Kursinnehåll

- *En översikt över språkteknologi:* delområden, tillämpningar och exempel.
- *Korpuslingvistik:* Reguljära uttryck, ändliga automater, introduktion till Python, konkordanser, tokenisering, frekvenslistor, kollokationer.
- *Morfologi och ordklasstagning:* Morfologi, transduktorer, ordklasstagning.
- *Frasstrukturgrammatiker:* Konstituent, syntaxträd, DCG-regler, unifiering.
- *Partiell parsning:* Detektering av multiord, detektering av substantivgrupp och verbgrupp, informationsextrahering, utvärdering.
- *Syntaktiska formalismer:* Konstituens och dependens, funktioner. Parsning, statistisk parsning, dependensparsning.
- *Semantik:* Formell semantik, lambda-kalkyl, lexikal semantik, predikat-argumentstruktur, ramsemantik, semantisk parsning.
- *Diskurs och dialog:* Referens och koreferens, diskurs och retorik, diskursrelationer, analys av diskursrelationer, dialog, talhandling, multimodalitet.

Kursens examination

Betygsskala: TH - (U,3,4,5) - (Underkänd, Tre, Fyra, Fem)

Prestationsbedömning: För betyg 3 krävs godkända inlämningsuppgifter (laborationer och individuella rapporter). För möjlighet till högre betyg krävs även skriftlig tentamen. Inlämningsuppgifterna (laborationerna) skall utföras i grupper om två studenter, men kan även göras individuellt. Den första laborationen ägnas åt de programmeringsverktyg som används i kursen. De sex övriga uppgifterna behandlas under de följande laborationstillfällena.

Om så krävs för att en student med varaktig funktionsnedsättning ska ges ett likvärdigt examinationsalternativ jämfört med en student utan funktionsnedsättning, så kan examinator efter samråd med universitetets avdelning för pedagogiskt stöd fatta beslut om alternativ examinationsform för berörd student.

Delmoment

Kod: 0113. **Benämning:** Statistiska tekniker för textanalys.

Antal högskolepoäng: 3,5. **Betygsskala:** UG. **Prestationsbedömning:** För godkänt betyg krävs fullgjorda laborationer. **Delmomentet omfattar:** Laborationer.

Kod: 0213. **Benämning:** Syntaktisk och semantisk textbehandling.

Antal högskolepoäng: 4. **Betygsskala:** UG. **Prestationsbedömning:** För godkänt betyg skall laborationerna vara fullgjorda. **Delmomentet omfattar:** Laborationer.

Kod: 0313. **Benämning:** Tentamen.

Antal högskolepoäng: 0. **Betygsskala:** TH. **Prestationsbedömning:** De som fullgör övriga delmoment på kursen får betyg 3. För dem som deltar i den frivilliga tentamen finns möjlighet att få högre betyg. För dem som tenderar med betyg 4 eller 5 som resultat blir detta slutbetyg på hela kursen. **Delmomentet omfattar:** Frivillig skriftlig tentamen.

Antagningsuppgifter

Förkunskapskrav:

- EDAA01 Programmeringsteknik - fördjupningskurs eller EDAA30 Programmering i Java - fortsättningskurs

Begränsat antal platser: Nej

Kursen överlappar följande kurser: EDA171

Kurslitteratur

- Pierre Nugues: Language Processing with Perl and Prolog, Theories, Implementation, and Application. Springer Verlag, 2014, ISBN: 978-3-642-41464-0.

Kontaktinfo och övrigt

Kursansvarig: Professor Pierre Nugues, Pierre.Nugues@cs.lth.se

Hemsida: <http://cs.lth.se/edan20>