



DT2119 Igenkänning av tal och talare 7,5 hp

Speech and Speaker Recognition

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

Fastställande

Kursplan för DT2119 gäller från och med VT18

Betygsskala

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

Utbildningsnivå

Avancerad nivå

Huvudområden

Datalogi och datateknik

Särskild behörighet

För fristående kursstuderande krävs 90 hp varav 45 hp inom matematik eller informationsteknik. Dessutom krävs engelska B eller motsvarande.

Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

Studenten ska efter genomgången kurs kunna

- använda de i kursen beskrivna metoderna för att känna igen tal eller talare
- konfigurera ett system till en given applikation
- anpassa och vidareutveckla befintliga system för tal- och talarigenkänning
- utvärdera system för tal- och talarigenkänning
- bedriva forskningsarbete inom området.

Kursinnehåll

Kursen består av föreläsningar, tre laborationer med inlämningsuppgifter, samt att skriva en uppsats i ett ämne valt i samråd med läraren. Uppsatsen presenteras dessutom muntligt under ett slutseminarium. Laborationerna består i att designa olika delar av en taligenkänningsapplikation, träna systemet och utvärdera dess prestanda.

Följande teoretiska delmoment ingår:

- algoritmer för träning, igenkänning samt adaptation till egenskaper hos talare och transmissionskanal, inklusive mönsterigenkänning, Hidden Markov Models (HMMs) och Deep Neural Networks (DNNs)
- metoder för att minska känsligheten mot störningar och avvikelser
- sannolighetsteori
- signalbehandling och parameterextraktion
- akustisk modellering av talljudens statistiska och tidsvarierande spektrala egenskaper
- statistisk modellering av språkbruk i spontant och formellt tal
- sökstrategier - grundläggande metoder och strategier för stora vokabulärer
- specifika analys- och beslutsmetoder för igenkänning av talare.

Dessutom ges viss praktisk inblick i att bygga en tillämpning. Här ingår att implementera vissa funktioner utifrån prototyper och att testa dem på riktig taldata.

Kurslitteratur

- Huang, X., Acero, A., Hon, H.-W. Spoken Language Processing – A Guide to Theory, Algorithm and System Development, Prentice Hall, 2001.
- Automatic Speech Recognition: A deep learning approach, Dong Yu and Li Deng, Springer 2015. You can download the PDF through KTH Library.
- Särtryck på artiklar inom talarigenkänning och andra ämnen som inte ingår i boken.

Examination

- LAB1 - Laboration, 4,5 hp, betygsskala: P, F

- PRO1 - Projekt, 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.

Övriga krav för slutbetyg

Laboration

Inlämningsuppgifter

Uppsats med presentation vid ett slutseminarium

Bedömning av två övriga kursdeltagares uppsatser och opposition på deras presentationer.

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.