

Text Mining

Programkurs

6 hp

Text Mining

TDDE16

Gäller från: 2021 VT

Fastställd av

Programnämnden för data- och
medieteknik, DM

Fastställandedatum

2020-09-29

Huvudområde

Informationsteknologi, Datateknik, Datavetenskap

Utbildningsnivå

Avancerad nivå

Fördjupningsnivå

A1X

Kursen ges för

- Masterprogram i datavetenskap
- Civilingenjör i datateknik
- Civilingenjör i informationsteknologi
- Civilingenjör i mjukvaruteknik
- Civilingenjör i teknisk fysik och elektroteknik - internationell
- Civilingenjör i teknisk fysik och elektroteknik

Rekommenderade förkunskaper

- diskret matematik, analys, linjär algebra, sannolikhetslära
- programmering, datastrukturer och algoritmer
- maskininläring

Kursens laborationsserie använder programspråket Python.

Lärandemål

Text mining utvecklar metoder för att komma åt information i och extrahera kunskap från stora mängder text. Denna kurs har som mål att ge praktisk erfarenhet av de huvudsakliga stegen i text mining: informationsökning, bearbetning av textdata, modellering, analys av experimentella resultat. Kursen avslutas med ett individuellt projektarbete där de studerande arbetar med en egendefinierad frågeställning.

Efter genomgången kurs ska den studerande kunna:

1. implementera text mining-metoder och tillämpa dem på realistiska problem
2. analysera och sammanfatta resultat från text mining-experiment
3. identifiera, formulera och lösa problem inom text mining-området
4. klart redogöra för och diskutera slutsatserna av sitt arbete

Kursinnehåll

- informationssökning
- grundläggande språkteknologiska metoder
- prediktiv modellering, särskilt textklassificering
- textklustring och temamodellering
- informationsextraktion
- valideringsmetoder
- tillämpningar av text mining
- programbibliotek och data

Undervisnings- och arbetsformer

Kursen ges i form av föreläsningar, datorlaborationer och handledning i samband med ett individuellt projektarbete.

Examination

| | | | |
|------|-------------------|------|------------|
| PRA1 | Projekt | 3 hp | U, 3, 4, 5 |
| LAB1 | Datorlaborationer | 3 hp | U, G |

Kursens slutbetyg är betyget på PRA1.

Betygsskala

Fyrgradig skala, LiU, U, 3, 4, 5

Övrig information

Påbyggnadskurser

Språkteknologi

Om undervisnings- och examinationsspråk

Undervisningsspråk visas på respektive kurstillfälle på fliken "Översikt".
Examinationsspråk relaterar till undervisningsspråk enligt nedan:

- Om undervisningsspråk är Svenska ges kursen i sin helhet eller till stora delar på svenska. Observera att även om undervisningsspråk är svenska kan delar av kursen ges på engelska. Examinationsspråk är svenska.
- Om undervisningsspråk är Svenska/Engelska kan kursen i sin helhet ges på engelska vid behov. Examinationsspråk är svenska eller engelska.
- Om undervisningsspråk är Engelska ges kursen i sin helhet på engelska. Examinationsspråk är engelska.

Övrigt

Kursen bedrivs på ett sådant sätt att både mäns och kvinnors erfarenhet och kunskaper synliggörs och utvecklas.

Planering och genomförande av kurs skall utgå från kursplanens formuleringar. Den kursvärdering som ingår i kursen skall därför genomföras med kursplanen som utgångspunkt.

Institution

Institutionen för datavetenskap

Studierektor eller motsvarande

Peter Dalenius

Examinator

Marco Kuhlmann

Kurshemsida och andra länkar

<http://www.ida.liu.se/~TDDE16/>

Undervisningstid

Preliminär schemalagd tid: 28 h

Rekommenderad självstudietid: 132 h

Generella bestämmelser

Kursplan

För varje kurs ska en kursplan finnas. I kursplanen anges kursens mål och innehåll samt de särskilda förkunskaper som erfordras för att den studerande skall kunna tillgodogöra sig undervisningen.

Schemaläggning

Schemaläggning av kurser görs enligt, för kursen, beslutad blockindelning.

Avbrott på kurs

Enligt rektors beslut om regler för registrering, avregistrering samt resultatrapportering (Dnr LiU-2015-01241) skall avbrott i studier registreras i Ladok. Alla studenter som inte deltar i kurs man registrerat sig på är alltså skyldiga att anmäla avbrottet så att kursregistreringen kan tas bort. Avanmälan från kurs görs via webbformulär, www.lith.liu.se/for-studenter/kurskomplettering?l=sv.

Inställd kurs

Kurser med få deltagare (< 10) kan ställas in eller organiseras på annat sätt än vad som är angivet i kursplanen. Om kurs skall ställas in eller avvikelser från kursplanen skall ske prövas och beslutas detta av dekanus.

Riktlinjer rörande examination och examinator

Se Beslut om Riktlinjer för utbildning och examination på grundnivå och avancerad nivå vid Linköpings universitet Dnr LiU-2019-00920, (<http://styrdokument.liu.se/Regelsamling/VisaBeslut/917592>).

Examinator för en kurs ska inneha en läraranställning vid LiU i enlighet med LiUs anställningsordning, Dnr LiU-2017-03931 (<https://styrdokument.liu.se/Regelsamling/VisaBeslut/622784>). För kurser på avancerad nivå kan följande lärare vara examinator: professor (även adjungerad och gästprofessor), biträdande professor (även adjungerad), universitetslektor (även adjungerad och gästlektor), biträdande universitetslektor eller postdoktor. För kurser på grundnivå kan följande lärare vara examinator: professor (även adjungerad och gästprofessor), biträdande professor (även adjungerad), universitetslektor (även adjungerad och gästlektor), biträdande universitetslektor, universitetsadjunkt (även adjungerad och gästadjunkt) eller postdoktor. I undantagsfall kan även en Timlärare utses som examinator på både grund- och avancerad nivå, se Tekniska fakultetsstyrelsen vidaredelegationer.

Examination

Principer för tentamina

Skriftlig och muntlig tentamen samt digital salstentamen och datortentamen ges minst tre gånger årligen; en gång omedelbart efter kursens slut, en gång i augustiperioden samt vanligtvis i en av omtentamensperioderna. Annan placering beslutas av programnämnden.

Principer för tentamensschemat för kurser som följer läsperioderna:

- kurser som ges Vt1 förstagångstentteras i mars och omtentteras i juni och i augusti
- kurser som ges Vt2 förstagångstentteras i maj och omtentteras i augusti och i oktober
- kurser som ges Ht1 förstagångstentteras i oktober och omtentteras i januari och augusti
- kurser som ges Ht2 förstagångstentteras i januari och omtentteras i mars och i augusti

Tentamensschemat utgår från blockindelningen men avvikelser kan förekomma främst för kurser som samläses/samtentteras av flera program samt i lägre årskurs.

För kurser som av programnämnden beslutats vara vartannatårskurser ges tentamina 3 gånger endast under det år kursen ges.

För kurser som flyttas eller ställs in så att de ej ges under något eller några år ges tentamina 3 gånger under det närmast följande året med tentamenstillfällena motsvarande dem som gällde före flyttningen av kursen.

När en kurs ges för sista gången ska ordinarie tentamen och två omtentamina erbjudas. Därefter fasas examinationen ut med tre tentamina samtidigt som tentamen ges i eventuell ersättningskurs under det följande läsåret. Om ingen ersättningskurs finns ges tre tentamina i omtentamensperioder under det följande läsåret. Annan placering beslutas av programnämnden. I samtliga fall ges dessutom tentamen ytterligare en gång under det därpå följande året om inte programnämnden föreskriver annat.

Om en kurs ges i flera perioder under året (för program eller vid skilda tillfällen för olika program) beslutar programnämnden/programnämnderna gemensamt om placeringen av och antalet omtentamina.

Omprov övriga examinerande moment

För riktlinjer för omprov vid andra examinerande moment än skriftliga tentamina, digital salstentamina och datortentamina hänvisas till de generella LiU-riktlinjerna för examination och examinator, <http://styrdokument.liu.se/Regelsamling/VisaBeslut/917592>.

Anmälan till tentamen

Fram till 31 januari 2021 gäller enligt tidigare riktlinjer: För deltagande i skriftlig tentamen, digital salstentamen och datortentamen krävs att den studerande gjort förhandsanmälan i Studentportalen under anmälningsperioden, dvs tidigast 30

dagar och senast 10 dagar före tentamensdagen. Anvisad sal meddelas fyra dagar före tentamensdagen via e-post. Studerande, som inte förhandsanmält sitt deltagande riskerar att avvisas om plats inte finns inom ramen för tillgängliga skrivningsplatser.

Från 1 februari 2021 gäller nya riktlinjer för anmälan till skriftliga salsskrivningar inklusive digitala tentamina i sal, Dnr LiU-2020-02033 (se beslut i regelsamlingen <https://styrdokument.liu.se/Regelsamling/VisaBeslut/622682>).

Teckenförklaring till tentaansmälningssystemet:

** markerar att tentan ges för näst sista gången

* markerar att tentan ges för sista gången

Ordningsföreskrifter för studerande vid tentamensskrivningar

Se särskilt beslut i

regelsamlingen: <http://styrdokument.liu.se/Regelsamling/VisaBeslut/622682>

Plussning

Vid Tekniska högskolan vid LiU har studerande rätt att genomgå förnyad examination (s.k. plussning) för högre betyg på skriftliga tentamina, digitala salstentamina och datortentamina, dvs samtliga provmoment med modulkod TEN, DIT och DAT. På övriga examinationsmoment ges inte möjlighet till plussning, om inget annat anges i kursplan.

Plussning är ej möjlig på kurser som ingår i utfärdad examen.

Betyg och examinationsformer

Företrädesvis skall betygen underkänd (U), godkänd (3), icke utan beröm godkänd (4) och med beröm godkänd (5) användas.

- Kurser med skriftlig tentamen och digital salstentamen skall ge betygen (U, 3, 4, 5).
- Kurser med stor del tillämpningsinriktade moment såsom laborationer, projekt eller grupparbeten får ges betygen underkänd (U) eller godkänd (G).
- Examensarbete samt självständigt arbete ger betyg underkänd (U) eller godkänd (G).

Examinationsmoment och modulcoder

Nedan anges vad som gäller för de examinationsmoment med tillhörande modulcod som tillämpas vid Tekniska fakulteten vid Linköpings universitet.

- Skriftlig tentamen (TEN) och digital salstentamen (DIT) skall ge betyg (U, 3, 4, 5).
- Examinationsmoment som kan ge betygen underkänd (U) eller godkänd (G) är laboration (LAB), projekt (PRA), kontrollskrivning (KTR), digital kontrollskrivning (DIK), muntlig tentamen (MUN), datortentamen (DAT), uppgift (UPG), hemtentamen (HEM).
- Övriga examinationsmoment där examinationen uppfylls framför allt

genom aktivt deltagande som basgrupp (BAS) eller moment (MOM) ger betygen underkänd (U) eller godkänd (G).

- Examinationsmomenten Opposition (OPPO) och Auskultation (AUSK) inom examensarbetet ger betyg underkänd (U) eller godkänd (G).

Allmänt gäller att:

- Obligatoriska kursmoment skall vara poängsatta och ges en modulkod.
- Examinationsmoment som ej är poängsatt får ej vara obligatoriskt. Det är frivilligt att delta på dessa moment och information om det samt tillhörande villkor skall tydligt framgå i den beskrivande texten.
- För kurser med flera examinationsmoment med graderad betygsskala skall det anges hur slutbetyg på kursen vägs samman.

För obligatoriska moment gäller att: Om det finns särskilda skäl, och om det med hänsyn till det obligatoriska momentets karaktär är möjligt, får examinator besluta att ersätta det obligatoriska momentet med en annan likvärdig uppgift. (I enlighet med LiU-riktlinjerna

<http://styrdokument.liu.se/Regelsamling/VisaBeslut/917592>).

För samtliga examinationsmoment gäller att: Om LiU:s koordinator för studenter med funktionsnedsättning har beviljat en student rätt till anpassad examination vid salstentamen har studenten rätt till det. Om koordinatören istället har gett studenten en rekommendation om anpassad examination eller alternativ examinationsform, får examinator besluta om detta om examinator bedömer det möjligt utifrån kursens mål. (I enlighet med LiU-riktlinjerna

<http://styrdokument.liu.se/Regelsamling/VisaBeslut/917592>).

Rapportering av examinationsresultat

Rapportering av den studerandes examinationsresultat sker på respektive institution.

Plagiering

Vid examination som innebär rapportskrivande och där studenten kan antas ha tillgång till andras källor (exempelvis vid självständiga arbeten, uppsatser etc) måste inlämnat material utformas i enlighet med god sed för källhänvisning (referenser eller citat med angivande av källa) vad gäller användning av andras text, bilder, idéer, data etc. Det ska även framgå ifall författaren återbrukat egen text, bilder, idéer, data etc från tidigare genomförd examination, exempelvis från kandidatarbete, projektrapporter etc. (ibland kallat självplagiering).

Underlåtelse att ange sådana källor kan betraktas som försök till vilseledande vid examination.

Försök till vilseledande

Vid grundad misstanke om att en student försökt vilseleda vid examination eller när en studieprestation ska bedömas ska enligt Högskoleförordningens 10 kapitel examinator anmäla det vidare till universitetets disciplinnämnd. Möjliga konsekvenser för den studerande är en avstängning från studierna eller en

varning. För mer information
se <https://www.student.liu.se/studenttjanster/lagar-regler-rattigheter?l=sv>.

Regler

Universitetet är en statlig myndighet vars verksamhet regleras av lagar och förordningar, exempelvis Högskolelagen och Högskoleförordningen. Förutom lagar och förordningar styrs verksamheten av ett antal styrdokument. I Linköpings universitets egna regelverk samlas gällande beslut av regelkaraktär som fattats av universitetsstyrelse, rektor samt fakultets- och områdesstyrelser.

LiU:s regelsamling angående utbildning på grund- och avancerad nivå nås på http://stydokument.liu.se/Regelsamling/Innehall/Utbildning_pa_grund_och_avancerad_niva.