


[Hjälp](#)

Försättsida för: TDDDD01/TEN1 går den 2013-06-03 (14-18) i Linköping.

Byt till... ▼

Välj

Ändra lösenord ▼

Välj

[\[Se bokade lokaler | Se beställningen \]](#)

Fyll i uppgifter för försättsbladet

Datum för tentamen	2013-06-03
Sal (1) Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	TER1
Tid	14-18
Kurskod	TDDDD01
Provkod	TEN1
Kursnamn/benämning Provnamn/benämning	Språkteknologi Skriftlig tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	<input type="text" value="11"/>
Jour/Kursansvarig Ange vem som besöker salen	<input type="text" value="Lars Ahrenberg"/>
Telefon under skrivtiden	<input type="text" value="013-282422"/>
Besöker salen ca kl.	<input type="text" value="14.45"/>
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	<input type="text" value="Liselotte Lundberg 281278 liselotte.lundberg"/>
Tillåtna hjälpmedel	<input type="text" value="Inga"/>
Övrigt	<input type="text"/>
Vilken typ av papper ska användas, rutigt eller linjerat	<input type="text" value="Valfritt"/>
Antal exemplar i påsen	<input type="text"/>

Välj språk:

- Svenska
- Engelska

[Se resultatet](#)

Informationsansvarig: [Marie Plantin](#)

Senast ändrad: Tue May 28 14:42:01 2013

Denna sidas del-URL: [/tentaanm/front_page.pl](#)

Teknisk ansvarig: [UNIT](#)

TENTAMEN

TDDD01 Språkteknologi
måndag 3 juni 2013 kl. 14-18

Inga hjälpmedel är tillåtna.

Maximal poäng finns angiven för varje fråga. Maximal poäng på hela tentan är 26. 13 poäng ger säkert godkänt (betyg 3). Planerade betygsgränser: 4 – 18; 5 – 22.

Besvara alla frågor. Det går bra att besvara flera frågor på samma papper.

1. Vad skiljer lemmatisering från stemming? (2p)
2. (a) Vad menas med att en ändlig tillståndsautomat (eng. finite-state automaton) är deterministisk? (b) Ange en sådan automat som genererar samma strängmängd som det reguljära uttrycket 'b(a|c)b*' (2p)
3. I en korpus hittades 120 ord som slutar på 'bar', bl.a. *snobbar* och *sökbar*. Av dessa befanns 40 vara substantiv, 65 vara adjektiv och 15 vara verb. (a) Ange MLE-uppskattningen med användningen av denna korpus för sannolikheten att ett ord som slutar på 'bar' är ett substantiv; (b) Hur kan vi ändra denna uppskattning om vi vill ge utrymme för möjligheten att egennamn också skulle kunna sluta på 'bar', trots att inga sådana egennamn setts i korpusen? Svara med en beskrivning eller en lämplig omräkning. (3p)
4. Ett namnigenkänningsprogram kritiserades för att generera för många 'falska positiva'. (a) Vad betyder det? (b) Vilket eller vilka av måtten precision eller recall påverkas av antalet falska positiva? (2p)
5. Vid ordprediktion är man intresserad av sannolikheten $p(w_i | w_1 w_2 \dots w_{i-1})$. (a) Vad innebär det att använda en bigrammodell för denna sannolikhet? (b) Vilket är det vanligaste måttet för att mäta prestanda för en ordprediktionsmodell? (2p)
6. Vad är Levenshteinavståndet mellan orden *brottnig* och *boxning*? Motivera ditt svar. (2p)
7. Förklara vad som menas med partiell parsning (eller chunkning)? Visa sedan hur chunkning kan hanteras som ett taggningsproblem. (2p)

8. (a) Förklara begreppen anaför och antecedent med exempel. (b) Varför är det viktigt att kunna känna igen anaförer och antecedenter exempelvis i en tillämpning som informationsextraktion? (2p)
9. Modellen den brusiga kanalen (eng. *Noisy channel*) kan t.ex. tillämpas på problemet översättning med beslutsregeln "Givet en mening F på ett främmande språk, välj den engelska mening E* som har störst sannolikhet att ge upphov till F". Formulera beslutsregeln matematiskt och illustrera modellen grafiskt. (3p)
10. Relationer, som den mellan en vara och ett pris kan uttryckas på många sätt i naturligt språk, t.ex med uttryck som *X kostar Y, jag vill ha Y för X, priset för X är Y*. Beskriv en metod att hitta en stor mängd sådana relationsangivande uttryck i en större textkorpus. (3p)
11. I en chart-parser som gavs meningen *vanlig mjölk är godast* som indata fanns i ett visst läge bland annat följande tillstånd/bågar:
- $S \rightarrow . NP VP \quad [0,0]$
 $NP \rightarrow A . N \quad [0,1]$
- Ange två andra tillstånd/bågar som också måste finnas i charten. Motivera ditt svar. (3p)