



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2013-08-23
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	TDDD01
Provkod	TEN1
Kursnamn/benämning Provnamn/benämning	Språkteknologi Skriftlig tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	11
Jour/Kursansvarig Ange vem som besöker salen	Lars Ahrenberg
Telefon under skrivtiden	/2422
Besöker salen ca kl.	14.45
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Liselotte Lundberg, tel 127828, liselotte.lundberg@liu.se
Tillåtna hjälpmedel	inga
Övrigt	
Vilken typ av papper ska användas, rutigt eller linjerat	valfritt
Antal exemplar i påsen	

TENTAMEN

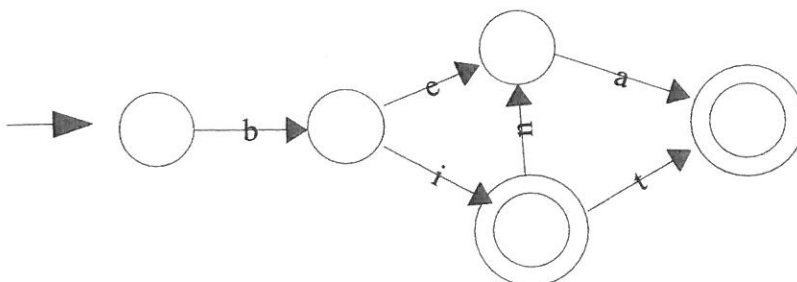
TDDD01 Språkteknologi
fredag 23 augusti 2013 kl. 14-18

Inga hjälpmedel är tillåtna.

Maximal poäng finns angiven för varje fråga. Maximal poäng på hela tentan är 25. 13 poäng ger säkert godkänt (betyg 3). Planerade betygsgränser: 4 – 17; 5 – 21.

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

1. Förklara vad som menas med att normalisera en text och ange tre olika textbearbetningar som kan ingå i textnormalisering. (3p)
2. Definiera måtten precision och recall och ange två språkteknologiska processer som utvärderas med dessa mått. (3p)
3. Vilket är Levenshteinavståndet mellan orden *broar* och *bryggor*? Svaret skall motiveras. (2p)
4. 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 trigrammodell för denna sannolikhet? (b) Vad menas med en trigrammodell med back-off? (2p)
5. Förklara vad som menas med en s.k. förväxlingsmatris (eng. *confusion matrix*) och ge något exempel på språkteknologisk användning av sådana matriser. (2p)
6. Förklara i ord vad w^* står för i det matematiska uttrycket $w^* = \underset{w}{\operatorname{argmax}} p(w|kan)$. (1p)
7. Vilka av följande teckensträngar definieras av den här automaten: be, ben, bi, bin, bina, bit, bita? (2p)



8. I standardarkitekturen för ett frågebesvarande system brukar man hitta en modul för frågeanalys. Förklara kort vad denna modul gör och hur dess utdata används av andra moduler i systemet. (3p)
9. Förklara vad som menas med en språkmodell och varför man i regel behöver tillämpa s.k. utjämning (eng. *smoothing*) när man skapar språkmodeller. (2p)
10. Vilka är standardkomponenterna i ett informationsutvinningsystem (eng. *information extraction*). (2p)
11. Syntaktisk analys av meningar i naturligt språk görs ofta med en s.k. chart-parser. Förklara följande chartparsningsbegrepp: (a) chart, (b) predicering (eng. *prediction*), (c) kombinerings (eng. *completion*). (3p)