

Tentamen Taal- en Spraaktechnologie

28 juni 2010 13.15-15.00 uur

1. Hieronder staan observatiewaarschijnlijkheden voor vier POS.

VB=verb, base form (*make, try*)

TO=infinite marker (*to*)

NN=(common) singular or mass noun (*time, world, work*)

PPSS=other nominative pronoun (*I, we, they*)

	I	want	to	race
VB	0	.0093	0	.00012
TO	0	0	.99	0
NN	0	.000054	0	.00057
PPSS	.37	0	0	0

- a) Waarom sommeren rijen en kolommen niet tot 1?

Hieronder staan de transitiekansen tussen de POS, <s> betekent begin van de zin.

	VB	TO	NN	PPSS
<s>	.019	.0043	.041	.067
VB	.0038	.035	.047	.0070
TO	.83	0	.00047	0
NN	.0040	.016	.087	.0045
PPSS	.23	.00079	.0012	.00014

- b) Wat is de totale kans om *I race* tegen te komen.

- c) Bereken de waarschijnlijkste bijbehorende POS combinatie.

2. Gegeven een minicorpus met drie zinnen:

ik houd van Marie

het is Marie van wie ik houd

van Sofie houd ik niet

Sofie is de vriendin van Marie niet

ik houd Marie niet

Hoe groot zijn $P(\text{houd}|\text{ik})$, $P(\text{niet}|\text{Marie})$, $P(\text{WW}|\text{VNW})$?

3. Wat is de orde van grootte van de winst in het aantal berekeningen van het forward-backward algoritme ten opzichte van een directe berekening van $P(O|\lambda)$? Wat is de basis van deze winst?
4. Langs welke stappen kun je een difoon database maken?
5. Waarom zijn harmonischen zelf niet van belang voor automatische spraakherkenning maar hun omhullende spectrum wel?
6. Het zinnetje "taal is bijzonder mooi" wordt door een ASH systeem herkend als "paar is blij van de mooi". Wat is de word error rate? En de foneem error rate?

7. Geef aan wat het verschil is tussen een deterministische en een non-deterministische FSA, gegeven een toestand en een alfabetsymbool.

Verschillen deterministische vs non-deterministische FSA's wat betreft de klasse van talen die door die machines worden herkend?

-Zo ja, welke soort machine herkent dan de grootste klasse van talen?

-Zo nee, verklaar waarom niet.

8. Een subklasse van de finite state transducers zijn de ***sequentiële*** transducers. Wat is de kenmerkende eigenschap van een sequentiële transducer? Waarom is deze subklasse van transducers van belang?

9. Maak een transducer die, gegeven de infinitief van een regelmatig werkwoord in het Nederlands, het voltooid deelwoord maakt.

Voorbeeld. INPUT: *werken*, OUTPUT: *gewerkt*

Gebruik de notatie van het practicum:

macro(voltdeelw, ...).

Bouw de transducer op als de ***compositie*** van een aantal eenvoudige stappen: uitgang weghalen, 't kofschip regel toepassen, (ver|ont)dubbeling van (mede)klinkers, enzovoort.

Je transducer moet werken voor de volgende gevallen:

haten, zweten, hakken, strekken, maken, straffen, keffen, krassen, grazen, matsen, kuchen, sturen, kapen, kappen, verven, zweven.