

P6.2 Synthese: Kontrollfragen

Uwe Reichel

7. Januar 2011

1 Abschnitt Überblick

1. Worin besteht der Unterschied zwischen *concept to speech* und *text to speech*?
2. Nenne 3 Möglichkeiten der Generierung des akustischen Signals.
3. Nenne die TTS-Module in der Standardarchitektur.
4. Nenne 2 Anwendungsgebiete der Sprachsynthese.
5. Nenne 3 gängige Sprachsynthesysteme.

2 Abschnitt Textnormalisierung

6. Aus welchen Komponenten setzt sich die Textnormalisierung zusammen?
7. Was versteht man unter *Tokenisierung*?
8. Gib 3 Beispiele von *Nicht-Standard-Wörtern*.
9. Woraus besteht ein *Endlicher Automat* (mit Skizze)?
10. Gib ein Beispiel, wo in der Synthese er Einsatz findet.
11. Woraus besteht ein *Finite-State-Transducer* (mit Skizze)?
12. Gib ein Beispiel, wo in der Synthese er Einsatz findet.

3 Abschnitt Part-of-Speech-Tagging

13. Wofür wird die Wortartinformation benötigt?

14. Nenne je ein POS-Inventar für das Deutsche und das Englische.
15. Worin bestehen die beiden Hauptschwierigkeiten beim POS-Tagging?
16. In welche drei Klassen lassen sich POS-Verfahren grob kategorisieren?
17. Was versteht man unter dem *Noisy-Channel-Modell*, und wie lässt es sich auf das POS-Tagging beziehen?
18. Welche Umformung und welche zwei vereinfachenden Annahmen werden zur Berechnung von $\hat{G} = \arg \max_G [P(G|W)]$ getroffen (W : Wortfolge, G : zugrundeliegende POS-Folge)?
19. Zeichne den Ausschnitt eines Hidden-Markov-Modells, wie es für das POS-Tagging Einsatz finden könnte.
20. Skizziere unter Verwendung eines Trellis-Ausschnitts die grundsätzliche Funktionsweise des Viterbi-Verfahrens.
21. Was versteht man unter dem Konzept der *Successor Variety*, und wie lässt es sich für das POS-Tagging nutzbar machen?

4 Abschnitt Graphem-Phonem-Konvertierung (G2P)

22. Nenne mit jeweils einem Beispiel drei Einflussfaktoren auf die Graphem-Phonem-Abbildung.
23. Was versteht man unter *Alinierung*?
24. Was versteht man unter der *Levenshtein-Distanz*?
25. Berechne die Levenshtein-Distanz zwischen *Gen* und *Gans* mittels dynamischer Programmierung. Verwende hierbei folgende Kostenfunktion: Nullsubstitution kostet 0, alle anderen Editieroperationen 1.
26. Beschreibe das *Table Lookup with Defaults*-Verfahren nach van den Bosch et al. (1993)
27. Nenne mit kurzer Begründung zwei Lernverfahren, die sich zur G2P-Konvertierung eignen (Stichwörter: Variablenwerte, (un)überwachtes Lernen).
28. Wie sind Graphem-Objekte, deren Eigenschaften, sowie die zugehörigen Phonem-Zielwerte in einem Entscheidungsbaum repräsentiert (mit Skizze)?

29. Welche drei Fälle werden beim rekursiven Aufbau des C4.5-Entscheidungsbaums unterschieden?
30. Definiere *Entropie* und *Bedingte Entropie* (Formeln nicht nötig).
31. Beschreibe eine Restriktionen bei der Zuweisung der Wortbetonung in Simplex-Wörtern.
32. Wie lautet die *Compound stress*-Regel der metrischen Phonologie? Gib ein Beispiel.

5 Abschnitt Prosodie

33. Was versteht man unter *prosodischer Struktur*?
34. Beschreibe ein Verfahren zur textbasierten Vorhersage prosodischer Phrasengrenzen (abgesehen von der Nutzung der Satzzeichen).
35. Was wird unter dem *neutralen Akzent* verstanden?
36. Nenne zwei Einflussfaktoren auf die Akzentuierung eines Worts.
37. Lege anhand zweier Intonationsmodelle dar, mit welchen drei Dichotomien sich solche Modelle charakterisieren lassen.
38. Beschreibe einen Ansatz zur Überführung von Tonsymbolen in konkrete F0-Werte.
39. Aus welchen Komponenten setzt sich das Fujisaki-Modell zusammen?
40. Was versteht man allgemein unter *Analyse durch Synthese*?
41. Nenne drei Einflussfaktoren auf die Dauern von Lautsegmenten.
42. Beschreibe ein Lautdauermodell Deiner Wahl ein wenig genauer.

6 Abschnitt Konkatenative Synthese

43. Stelle knapp die zwei Ansätze *klassische Diphon-Synthese* und *Unit-Selection* einander gegenüber hinsichtlich Umfang der benötigten Daten, notwendiger Signalmanipulationen sowie Verkettung der Einheiten.
44. Was versteht man unter einem *Diphon*, und warum eignet sich diese Einheit zur Synthese?

45. Skizziere knapp die Erstellung einer Diphondatenbank.
46. Wie lassen sich mit *TD-PSOLA* Dauer, F0 und Intensität manipulieren?
47. Worin besteht die Normierung der Diphondatenbank bei *MBROLA*?
48. Was versteht man bei der *Unit-Selection-Synthese* unter *Target-* und *Join-Kosten*?
49. Wie lässt sich die *Unit-Selection-Synthese* mit einem HMM modellieren?
50. Wie lässt sich Verständlichkeit und Natürlichkeit der Synthese evaluieren?