

Information Retrieval

Aufgabe 5

- (1a) Die Markovkette ist irreduzibel, aber nicht aperiodisch, da $ggT(2, 4) = 2$. Somit ist die Kette auch nicht ergodisch.
- (1b) Die Markovkette ist irreduzibel, aber nicht aperiodisch, da $ggT(3, 6) = 3$. Somit ist die Kette auch nicht ergodisch.
- (1c) Die Markovkette ist irreduzibel und aperiodisch, da $ggT(2, 3) = 1$ und $ggT(3, 5) = 1$. Die Kette ist ergodisch, da sie positiv rekurrent und homogen ist.
- (1d) Die Markovkette ist nicht irreduzibel und daher nicht ergodisch. Sie ist aperiodisch, da $ggT(2, 3) = 1$ und $ggT(3, 4) = 1$ und Knoten 3 auf sich verweist.

Aufgabe 6

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/10 & 2/10 & 7/10 & 0 \\ 0 & 1/10 & 6/10 & 3/10 \\ 0 & 0 & 1/10 & 9/10 \end{pmatrix}$$

- (2a) Die Markovkette ist irreduzibel und aperiodisch, da jeder Knoten auf sich selbst verweist. Zudem ist sie positiv rekurrent und homogen.
- (2b)

$$\begin{aligned} w_1 &= 1/2w_1 + 1/10w_2 \\ w_2 &= 1/2w_1 + 2/10w_2 + 1/10w_3 \\ w_3 &= 7/10w_2 + 6/10w_3 + 1/10w_4 \\ w_4 &= 3/10w_3 + 9/10w_4 \\ 1 &= w_1 + w_2 + w_3 + w_4 \end{aligned}$$

Als stationäre Zustandswahrscheinlichkeiten erhalten wir $\pi = \begin{pmatrix} 1/146 \\ 5/146 \\ 35/146 \\ 105/146 \end{pmatrix} = \begin{pmatrix} 0,0068 \\ 0,0342 \\ 0,2397 \\ 0,7192 \end{pmatrix}$.

Aufgabe 7

Für die Entropie von Polynesich ergibt sich

$$\begin{aligned} H(P) &= - \sum_{i \in \{p,t,k,a,i,u\}} P(i) \log P(i) \\ &= -(4 \times \frac{1}{8} \log \frac{1}{8} + 2 \times \frac{1}{4} \log \frac{1}{4}) \\ &= 2 \frac{1}{2} \text{bits} \end{aligned}$$

Eine entropiebasierte Codierung verwendet also im Durchschnitt 2,5 Zeichen.

<i>p</i>	<i>t</i>	<i>k</i>	<i>a</i>	<i>i</i>	<i>u</i>
100	00	101	01	110	111