Technische Universität Kaiserslautern

ICSY Integrated Communication Systems

http://www.icsy.de

Multimedia Systems

WS 2009/2010

Exercise Course, January 19, 2010

Dipl.-Inf. Simon Schwantzer
University of Kaiserslautern, Germany
Integrated Communication Systems Lab

Email: schwantzer@informatik.uni-kl.de

Exercise 5.1

1. Let t be a text with a „normal“ 8-bit coding (ISO 8859-1).
   a) What is the entropy of a text which contains all 256 possible characters with the same probability?
      - Uniform distribution → maximal entropy → 8 bit
      - Formal:

$$H = \sum_i p_i \log_2 \frac{1}{p_i}[bit] = \sum_{i=0}^{255} \frac{1}{256} \log_2 \frac{1}{\frac{1}{256}}[bit] = \sum_{i=0}^{255} \frac{1}{256} \log_2 256[bit]$$

$$= \sum_{i=0}^{255} \frac{1}{256} * 8[bit] = 256 * \frac{1}{256} * 8[bit] = 8[bit]$$

Exercise 5.1

b) Determine an upper and a lower bound of the entropy for a text coded this way.

- Minimal entropy:
  - Source contains only one character.
  - 0 bit
- Maximal entropy:
  - Uniform distribution
  - 8 bit

Exercise 5.2

2. Let t be the following text:
   DIES IST EIN KURZER TEXT
   a) What are the absolute and relative occurrences of each character (including the spaces)?
      - Absolute occurrence = Amount of a character in the text.
      - Relative occurrence = Absolute occurrence / Total length

Char	Space	E	I	T	R	S	D	K	N	U	X	Z
Absolute	4	4	3	3	2	2	1	1	1	1	1	1
Relative	16,7%	16,7%	12,5%	12,5%	8,3%	8,3%	4,2%	4,2%	4,2%	4,2%	4,2%	4,2%

Exercise 5.2

2. Let t be the following text:
   DIES IST EIN KURZER TEXT

   b) What is the self information (in bit) of each character?

$$H_i = \log_2 \frac{1}{p_i}[bit] = -\log_2 p_i [bit] = -\frac{\ln p_i}{\ln 2}[bit]$$

Char	Space	E	I	T	R	S	D	K	N	U	X	Z
P_i	16,7%	16,7%	12,5%	12,5%	8,3%	8,3%	4,2%	4,2%	4,2%	4,2%	4,2%	4,2%
H_i	2,585	2,585	3,000	3,000	3,585	3,585	4,585	4,585	4,585	4,585	4,585	4,585

   c) What is the entropy (in bit) of each character?

$$H = \sum_i p_i \log_2 \frac{1}{p_i}[bit] = \sum_i p_i H_i \approx 3{,}355$$

Exercise 5.2

2. Let t be the following text:
   DIES IST EIN KURZER TEXT
   d) Based on the Shannon coding theorem, what do we know about the possible average code length?

- Entropy: $H = \sum_i p_i \log_2 \frac{1}{p_i} [bit] = \sum_i p_i H_i \approx 3{,}355$
- Shannon: $H \leq \bar{l} < H + 1$

→ There exists an entropy coding with an average code world length of between 3.355 and 4.355.

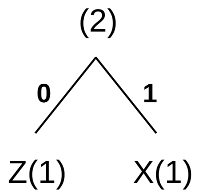
Exercise 5.2

2. Let t be the following text:
   DIES IST EIN KURZER TEXT
   d) Create a Huffman tree for t. How can the characters of t be encoded an decoded with this tree?

Char	Space	E	I	T	R	S	D	K	N	U	X	Z
Absolute	4	4	3	3	2	2	1	1	1	1	1	1
Relative	16,7%	16,7%	12,5%	12,5%	8,3%	8,3%	4,2%	4,2%	4,2%	4,2%	4,2%	4,2%

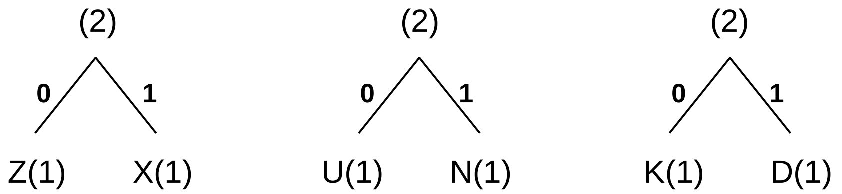
Exercise 5.2

Space	E	I	T	R	S	Z/X	D	K	N	U
16,7%	16,7%	12,5%	12,5%	8,3%	8,3%	8,3%	4,2%	4,2%	4,2%	4,2%



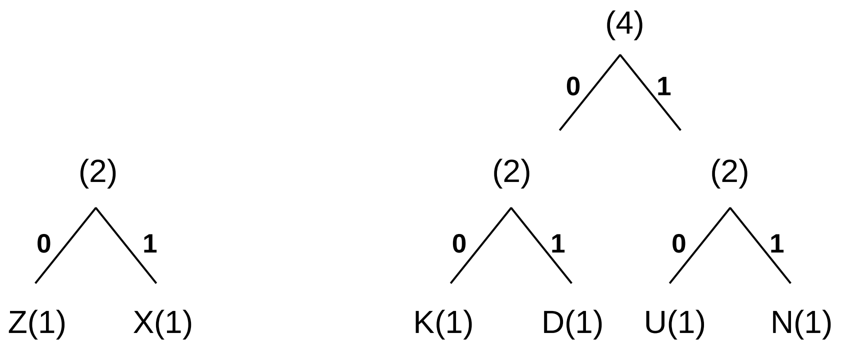
Exercise 5.2

Space	E	I	T	R	S	Z/X	U/N	K/D
16,7%	16,7%	12,5%	12,5%	8,3%	8,3%	8,3%	8,3%	8,3%



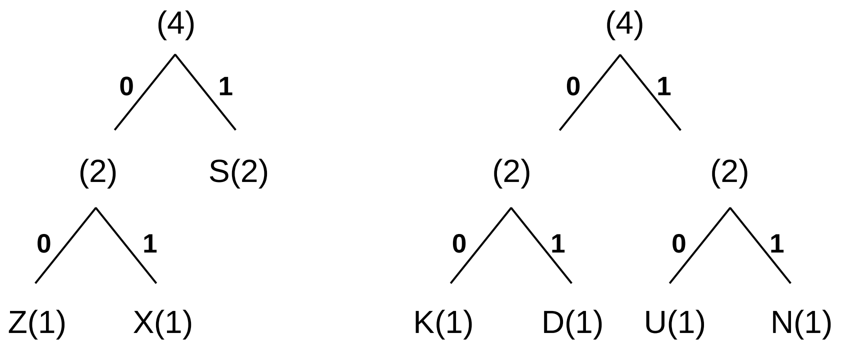
Exercise 5.2

Space	E	K/D/U/N	I	T	R	S	Z/X
16,7%	16,7%	16,7%	12,5%	12,5%	8,3%	8,3%	8,3%



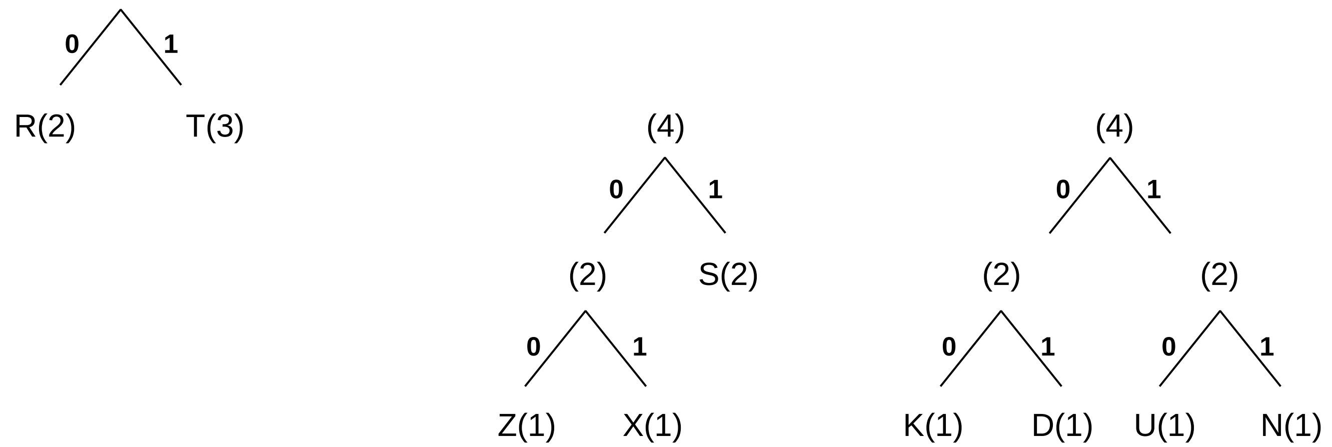
Exercise 5.2

Space	E	K/D/U/N	Z/X/S	I	T	R
16,7%	16,7%	16,7%	16,7%	12,5%	12,5%	8,3%



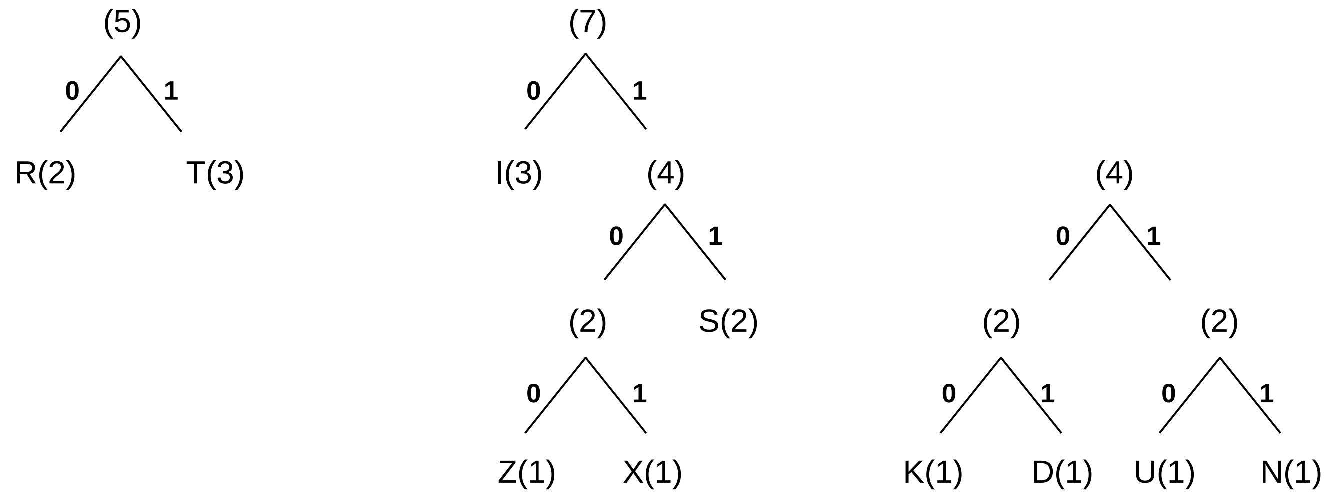
Exercise 5.2

R/T	Space	E	K/D/U/N	Z/X/S	I
20,8%	16,7%	16,7%	16,7%	16,7%	12,5%



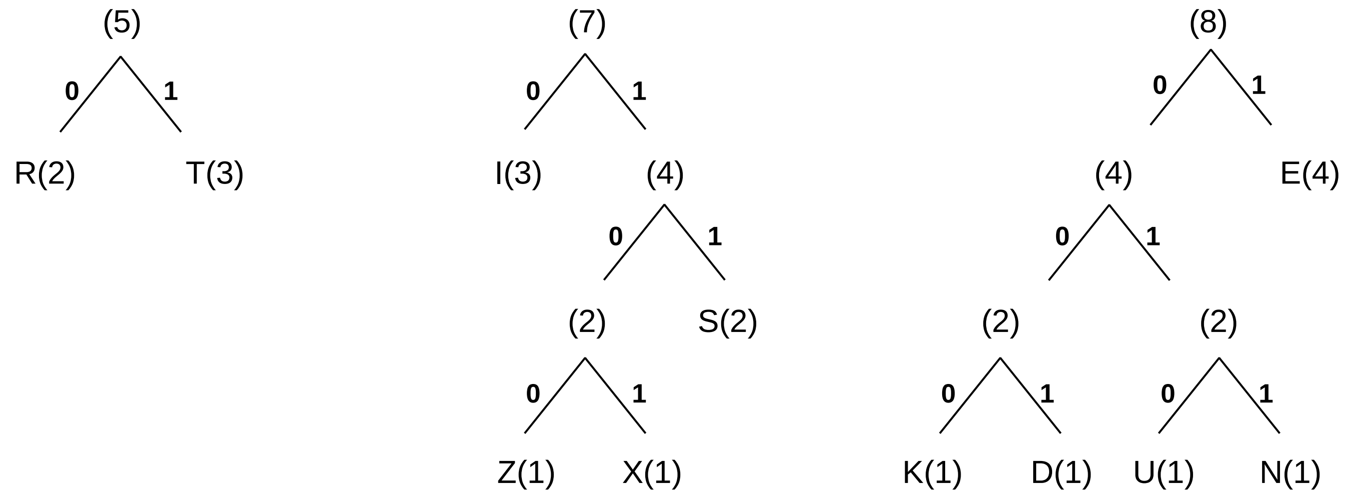
Exercise 5.2

I/Z/X/S	R/T	Space	E	K/D/U/N
29,2%	20,8%	16,7%	16,7%	16,7%



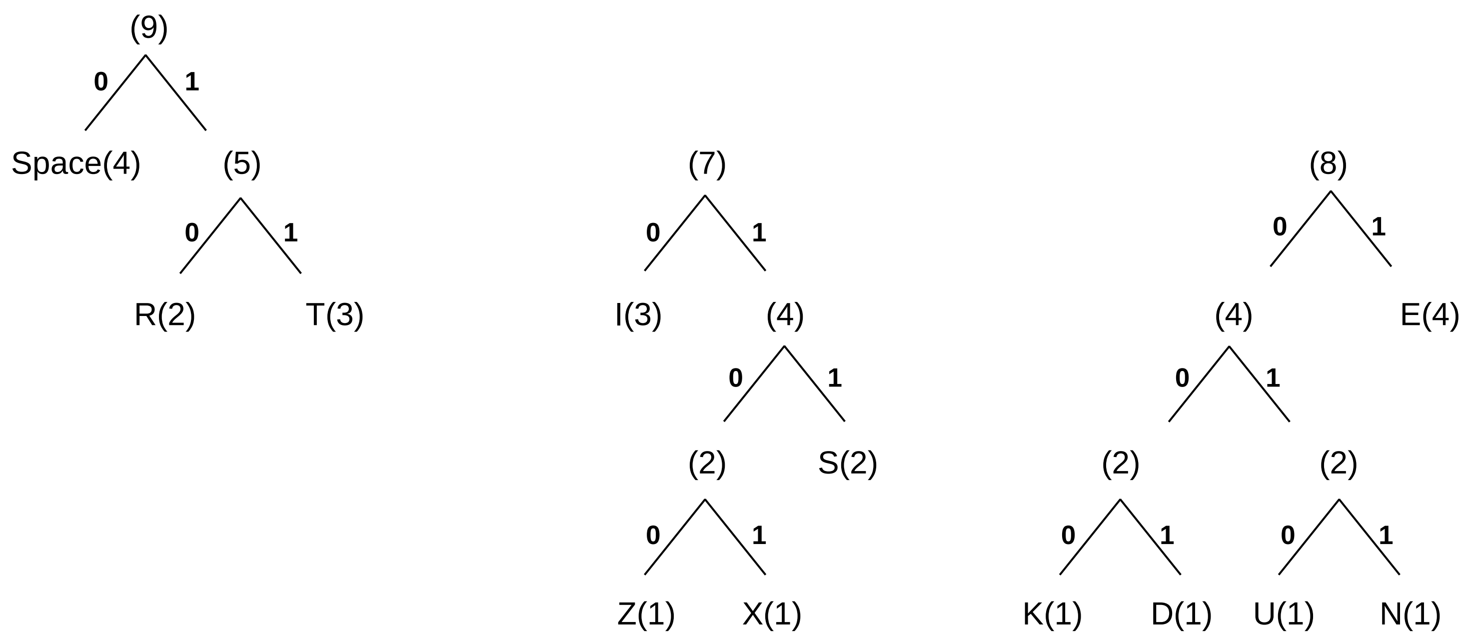
Exercise 5.2

K/D/U/N/E	I/Z/X/S	R/T	Space
33,3%	29,2%	20,8%	16,7%



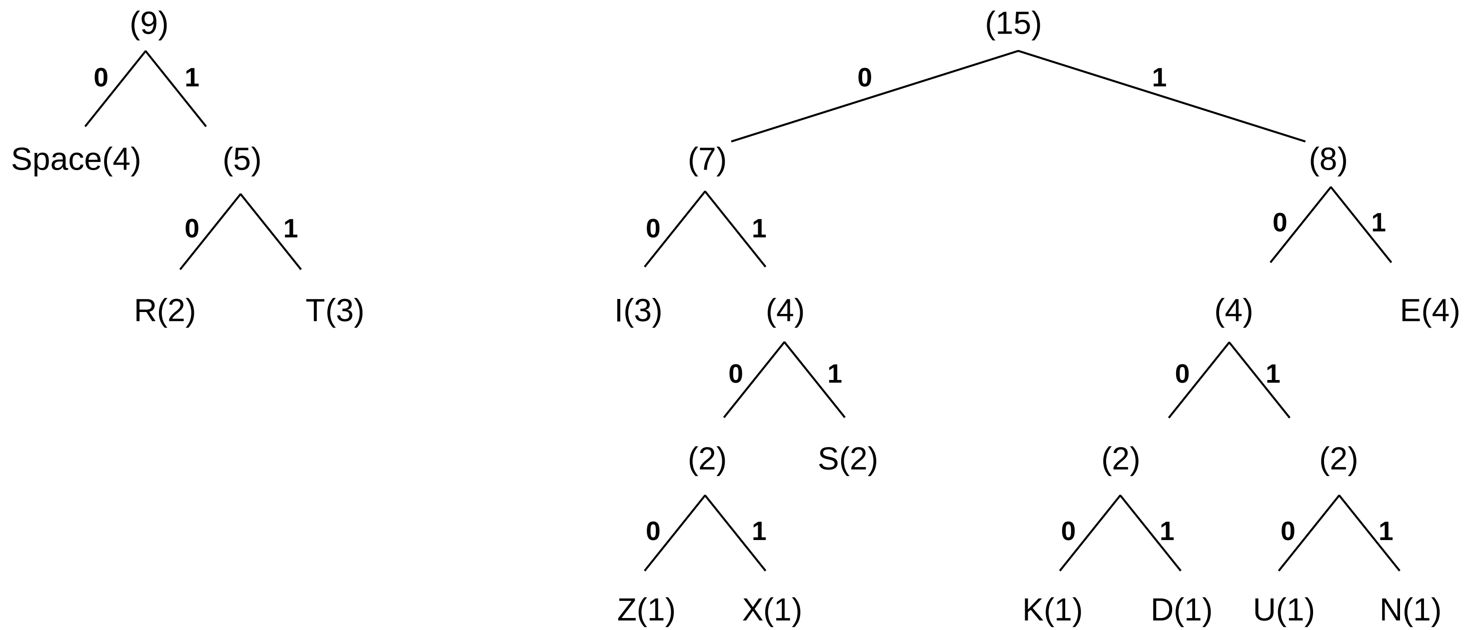
Exercise 5.2

Space/R/T	K/D/U/N/E	I/Z/X/S
37,5%	33,3%	29,2%



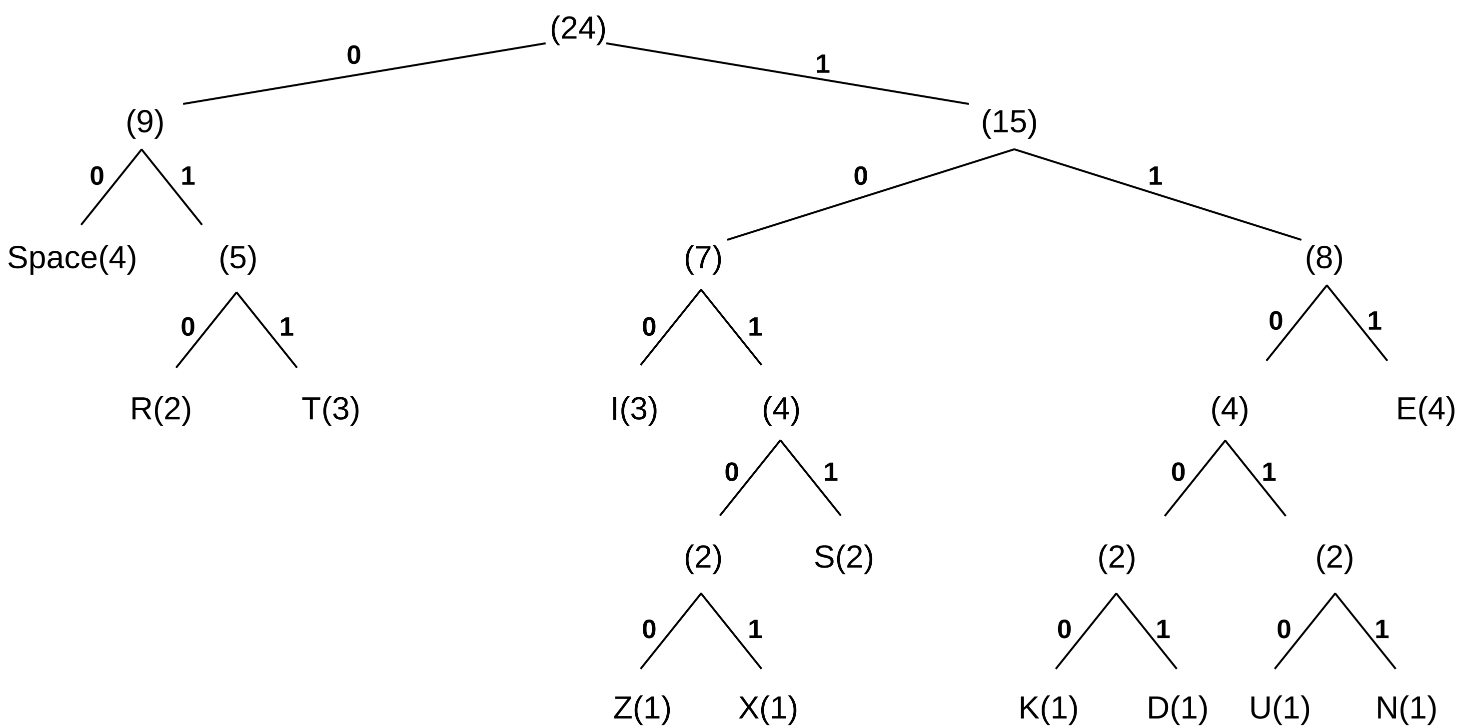
Exercise 5.2

I/Z/X/S/K/D/U/N/E	Space/R/T
62,5%	37,5%



Exercise 5.2

Space/R/T/I/Z/X/S/K/D/U/N/E
100,0%



Exercise 5.2

2. Let t be the following text:
   DIES IST EIN KURZER TEXT
   f) Determine the total length and the average codeword length of t (in bit) with
      i. a “normal” 8-bit coding (ISO 8859-1).
      ii. a coding fitted to the alphabet used in the text.
      iii. a Huffman
      iv. an (theoretically) optimal coding.

	“normal”	reduced	Huffman	optimal
Total length	192 Bit	96 Bit	82 Bit	80,529 Bit
Avg. codeword length	8 Bit	4 Bit	3,417 Bit	3,355 Bit

Notes

- Is the Huffman entropy coding optimal?
  - Average codeword lengths in the example:

    3.355 < 3.417 < 4.355

  → Often called "optimal" because it satisfies Shannon coding theorem.

  → Huffman does not necessary create an optimal average codeword length.
- When is a Huffman coding optimal?
  - A Huffman coding is optimal when all characters have an integral self information, i.e. all characters have an relative occurrence of

    $$P_i = \frac{1}{2^n}, n \in IN$$

    - n is the self information of the character.
    - Example: ABABACAD

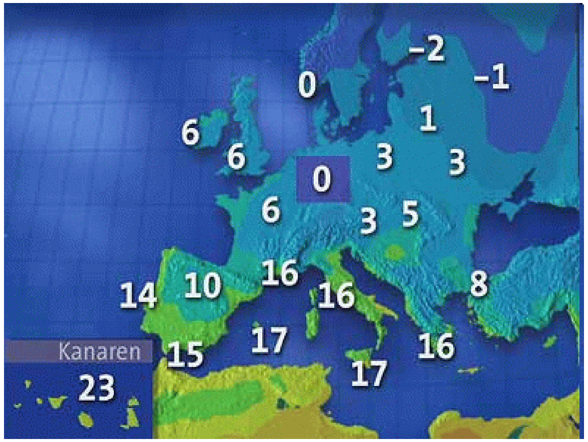
Statistically dependent information events

1. Consider the probability for the appearance of the letter "h" in a german text:
Is the probability of the letter "h" independent of prior symbols/letters?

2. Consider a drawing:

Is the probability of a green pixel independent of its neighbor pixel?

3. Consider a film:



Is the probability of a pixel independent of the same pixel in prior or following frames?