Compression
(SNLP tutorial 4)

Vilém Zouhar, Awantee Deshpande, Julius Steuer

TODOth, TODOth May 2021
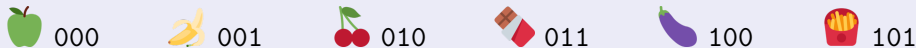
Organisation

TODO

Compression

TODO

Encoding

Task

Create coding (into binary) for the following recipe:

`apple apple banana cherries apple dark_chocolate eggplant banana cherries banana ...`

Fixed-width encoding

000   001   010   011   100   101

Length $= 14 \times 3 = 42$

Issues?

- Encoding for and ?
- What do 110 and 111 mean?

Encoding - Huffman

4× 🍏 A   4× 🍌 B   2× 🍒 C   2× 🍆 E   1× 🍫 D   1× 🍟 F

Huffman Bonus

- When will the Huffman tree be balanced?
- How do we store the tree? Does the efficiency of this matter?
- Are there undefined sequences of bits when using Huffman encoding?
- Does the result of Huffman encoding depend on the text ordering?
  E.g. 🍏🍌🍌🍫 vs. 🍌🍌🍫🍏🍌
- Can there be two equally good Huffman encodings?

Resources

1. Twitter emojis