

Школа лингвистики, 2021-22 уч. год

Дискретная математика для лингвистов

Разбор задач про (не)однозначно декодируемые коды (25-27 ноября 2021 года)

В. В. Кочергин, А. В. Михайлович

Задача 1. Выяснить, является ли код C с кодирующим алфавитом $\{0, 1, 2\}$ однозначно декодируемым:

- (a) $C = \{01, 12, 012, 111, 0102, 10112, 01112\}$;
- (b) $C = \{01, 10, 210, 201, 0210, 011022, 221\}$;
- (c) $C = \{01, 10, 210, 201, 0210, 011022\}$.

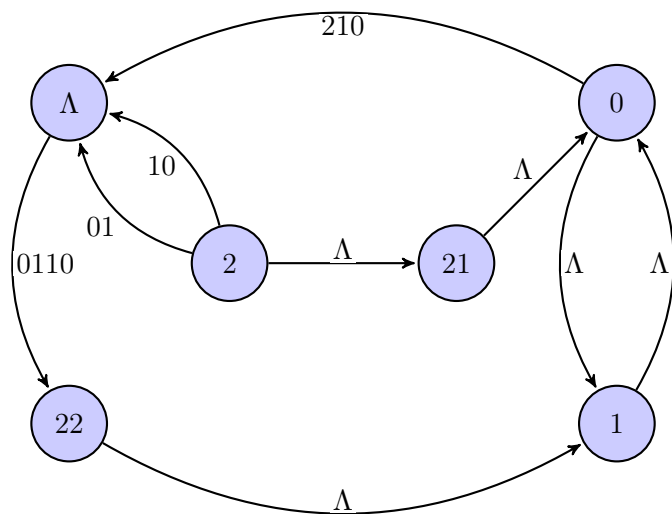
Если код префиксный или суффиксный, то код является разделимым. Если же код не префиксный или не суффиксный, то код может быть как разделимым ($\{0, 010\}$), так и неразделимым ($\{0, 1, 01\}$). Если код неразделимый, то найдётся слово, которое можно разбить на кодовые слова не единственным образом:

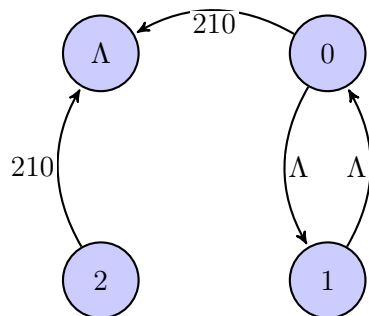
$$\underbrace{a_{i_1} a_{i_2} a_{i_3}} \underbrace{a_{i_4} \dots a_{i_r}} \underbrace{a_{i_{r+1}} \dots a_{i_s}} \underbrace{a_{i_{s+1}} \dots a_{i_t}} a_{i_{t+1}}, \dots, a_{i_p} \tag{1}$$

Чтобы проверить, возможно ли такое, сделаем следующее.

1. Находим подслова кодовых слов, которые одновременно являются и префиксом, и суффиксом кодовых слов, но сами не являются кодовыми словами. В слове 1 это могут быть $a_{i_3}, a_{i_4} \dots a_{i_r}, a_{i_{s+1}} \dots, a_{i_t}$ (если они не являются кодовыми словами). В задаче 1b — $\{0, 1, 2, 21, 22\}$, в задаче 1c — $\{0, 1, 2\}$.
2. Строим граф, в котором вершины помечены этими подсловами или пустым словом. Две вершины этого графа соединены направленным ребром, если есть кодовое слово, которое начинается с метки начала ребра, заканчивается на метке конца этого ребра, а в середине у него либо пустое слово, либо слово, полученное конкатенацией кодовых слов. При этом петли строим только от вершины, помеченной пустым словом и только в том случае, если петля помечается кодовым словом, которое разбивается на другие кодовые слова.

Построим графы для задач 1b и 1c.





Код является разделимым тогда и только тогда, когда в этом графе отсутствуют контуры и петли, проходящие через вершину Λ . Так, в графе задачи 1b цикл есть, то есть код не является разделимым (также по этому циклу можно найти неоднозначно декодируемое слово), а в графе задачи 1c — нет, то есть код является разделимым.

Задача 2. Выяснить, является ли код C с кодирующим алфавитом $\{0, 1, 2\}$ однозначно декодируемым:

- (a) $C = \{01, 12, 011, 01210, 20120, 2011220\}$;
- (b) $C = \{01, 12, 011, 01210, 201120, 2011220\}$;
- (c) $C = \{000, 0100, 10, 1001, 0010010\}$;
- (d) $C = \{01, 12, 01121, 21201\}$.

Задача 3. Является ли однозначно декодируемым код

- (a) $V = \{12, 001, 021, 102, 201, 0212, 00102\}$
- (b) $V = \{10, 01, 12, 012, 2100, 12010, 12011\}$
- (c) $V = \{12, 001, 021, 102, 201, 2012, 00102\}$