

# Алфавитное кодирование. Алгоритм распознавания однозначности алфавитного кодирования.

Селезнева Светлана Николаевна  
selezn@cs.msu.ru

факультет ВМК МГУ имени М.В. Ломоносова

Пусть заданы два конечных алфавита  $A$  и  $B$ .

Алфавит  $A$  назовем **исходным**, алфавит  $B$  — **кодирующим**.

**Кодированием** (из  $A$  в  $B$ ) называется произвольное отображение

$$\varphi : A^* \rightarrow B^* .$$

## Разделимость кодирования

Кодирование  $\varphi : A^* \rightarrow B^*$  называется **однозначным** (или **разделимым**), если для любых слов  $\alpha_1, \alpha_2 \in A^*$  из  $\alpha_1 \neq \alpha_2$  следует  $\varphi(\alpha_1) \neq \varphi(\alpha_2)$ .

Т.е. кодирование  $\varphi$  — разделимо, если **оно разным сообщениям сопоставляет различные коды**.

Другими словами, кодирование  $\varphi$  — однозначно, если **любое слово  $\beta \in B^*$  является кодом не более одного сообщения**.

# Алфавитное кодирование

Пусть  $A = \{a_1, \dots, a_r\}$  — исходный алфавит,  
 $B = \{b_1, \dots, b_q\}$  — кодирующий алфавит.

Кодирование  $\varphi : A^* \rightarrow B^*$  называется **алфавитным** (или **побуквенным**), если оно описывается следующей схемой:

1) заданы **различные** непустые коды букв алфавита  $A$ :

$$\begin{aligned}\varphi(a_1) &= B_1, B_1 \in B^*, \\ \varphi(a_2) &= B_2, B_2 \in B^*, \\ &\dots, \\ \varphi(a_r) &= B_r, B_r \in B^*,\end{aligned}$$

2) слова в алфавите  $A$  **кодируются побуквенно**, т.е. если  $\alpha \in A^*$ ,  $\alpha = a_{i_1} a_{i_2} \dots a_{i_m}$ , где  $m \geq 2$ , то

$$\varphi(\alpha) = \varphi(a_{i_1})\varphi(a_{i_2}) \dots \varphi(a_{i_m}) = B_{i_1} B_{i_2} \dots B_{i_m}.$$

# Алфавитный код

Пусть  $\varphi$  — алфавитное кодирование из  $A$  в  $B$ , т. е.

$$\varphi(a_1) = B_1, \varphi(a_2) = B_2, \dots, \varphi(a_r) = B_r.$$

Коды букв алфавита  $A$ , т. е. слова  $B_1, \dots, B_r$ , называются **кодowymi словами**.

Множество всех кодовых слов при кодировании  $\varphi$  назовем **алфавитным кодом**  $C_\varphi$ , т. е.

$$C_\varphi = \{B_1, \dots, B_r\}.$$

Алфавитный код  $C_\varphi$  назовем **однозначным** (или **разделимым**), если **кодирование  $\varphi$  — разделимо**.

Пусть  $C_\varphi = \{B_1, \dots, B_r\} \subseteq B^*$  — алфавитный код и  $\beta \in B^*$ .

Декодировать слово  $\beta$  означает **разбить его на последовательность кодовых слов** (если это возможно), т. е. представить в виде:

$$\beta = B_{i_1} B_{i_2} \dots B_{i_m},$$

где  $B_{i_1}, \dots, B_{i_m} \in C_\varphi$ .

Если код  $C_\varphi$  является разделимым, то для любого слова  $\beta \in B^*$  найдется **не более одного декодирования**.

# Граф разделимости алфавитного кода

Пусть  $C_\varphi = \{B_1, \dots, B_r\} \subseteq B^*$  — алфавитный код.

Построим *орграф*  $G_\varphi = (V_\varphi, E_\varphi)$  разделимости кода  $C_\varphi$ .

1. Множество вершин  $V_\varphi$ ,  $V_\varphi \subseteq B^*$ , состоит из пустого слова  $\Lambda$  и всех тех слов в алфавите  $B$ , которые **являются собственным префиксом** некоторого кодового слова и одновременно **собственным суффиксом** некоторого кодового слова (другого или, возможно, того же) и не являются никаким кодовым словом, т. е.

$$V_\varphi = \{\beta \in B^* \mid \begin{array}{l} 1) \exists B_i \in C_\varphi : B_i = \beta\beta', \beta' \neq \Lambda; \\ 2) \exists B_j \in C_\varphi : B_j = \beta''\beta, \beta'' \neq \Lambda; \\ 3) \beta \neq B_k, k = 1, \dots, r \}. \end{array}$$

# Граф разделимости алфавитного кода

Итак,  $C_\varphi = \{B_1, \dots, B_r\} \subseteq B^*$  — алфавитный код.

2. Опишем множество дуг  $E_\varphi$ : если  $\beta', \beta'' \in V_\varphi$ , то  $(\beta', \beta'') \in E_\varphi$ , если найдется такое кодовое слово  $B_i$  и такая последовательность  $D$  кодовых слов  $B_{i_1}, \dots, B_{i_k}$ , что

$$B_i = \beta' B_{i_1} \dots B_{i_k} \beta'',$$

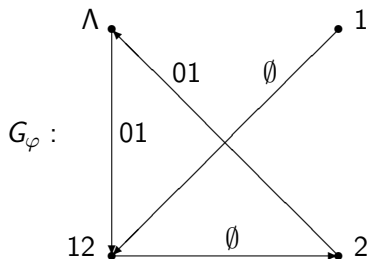
причем если  $\beta' = \beta'' = \Lambda$ , то  $k \geq 2$ ; если  $\beta' \neq \Lambda$  или  $\beta'' \neq \Lambda$ , то  $k \geq 1$ ; если  $\beta', \beta'' \neq \Lambda$ , то  $k \geq 0$ .

При этом дуге  $(\beta', \beta'') \in E_\varphi$  приписываем пометку  $D$ , где  $D = B_{i_1}, \dots, B_{i_k}$ .



## Граф разделимости алфавитного кода

**Пример.** Пусть  $C_\varphi = \{01, 201, 112, 122, 0112\}$ . Построим граф  $G_\varphi = (V_\varphi, E_\varphi)$ . Получаем:  $V_\varphi = \{\Lambda, 1, 2, 12\}$ .



# Критерий разделимости алфавитного кода

**Теорема.** Алфавитный код  $C_\varphi$  является разделимым тогда и только тогда, когда в графе  $G_\varphi$  отсутствуют ориентированные циклы (в том числе, и петли), проходящие через вершину  $\Lambda$ .

**Доказательство.** Пусть  $C_\varphi = \{B_1, \dots, B_r\} \subseteq B^*$  — алфавитный код и  $G_\varphi$  — граф разделимости кода  $C_\varphi$ .

# Критерий разделимости алфавитного кода

**Доказательство.** 1. Пусть код  $C_\varphi$  не является разделимым.

Значит, найдется слово  $\beta \in B^*$  **наименьшей длины**, которое допускает не менее двух декодирований.

Пусть  $\beta = B'_1 B'_2 \dots B'_{t_1}$  — разбиение слова  $\beta$  на кодовые слова в 1-м декодировании и  $\beta = B''_1 B''_2 \dots B''_{t_2}$  — разбиение слова  $\beta$  на кодовые слова во 2-м декодировании.

Обозначим:  $l'_i = |B'_i|$ ,  $i = 1, \dots, t_1$ , и  $l''_i = |B''_i|$ ,  $i = 1, \dots, t_2$ .

Пусть, для определенности,  $l''_1 > l'_1$ .

## Критерий делимости алфавитного кода

Доказательство. Найдем такое число  $k_1$ , что

$$\sum_{i=1}^{k_1-1} l'_i < l''_1, \quad \sum_{i=1}^{k_1} l'_i > l''_1.$$

Заметим, что равенства здесь быть не может, т. к. **в этом случае слово  $\beta$  можно было бы уменьшить**, что не так.

Тогда  $B''_1 = B'_1 \dots B'_{k_1-1} \beta_1$  для некоторого слова  $\beta_1 \in B^*$ ,  $\beta_1 \neq \Lambda$ .

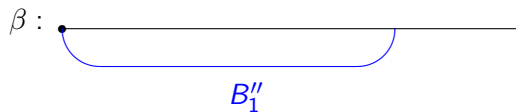
Отметим, что слово  $\beta_1$  является собственным префиксом кодового слова  $B'_{k_1}$  и собственным суффиксом кодового слова  $B''_1$ , а также **не является никаким кодовым словом**.

Значит, в графе  $G_\varphi$  присутствует дуга  $e_1 = (\Lambda, \beta_1) \in E_\varphi$ , которой приписана пометка  $D_1 = B'_1 \dots B'_{k_1-1}$ .

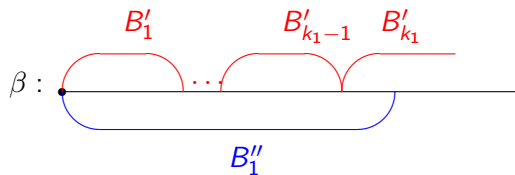
## Пояснение выбора числа $k_1$

$\beta$ : • \_\_\_\_\_

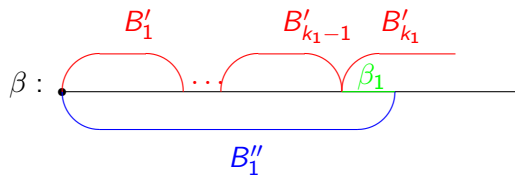
# Пояснение выбора числа $k_1$



# Пояснение выбора числа $k_1$



# Пояснение выбора числа $k_1$





## Критерий разделимости алфавитного кода

Доказательство. Теперь найдем такое число  $k_2$ , что

$$|\beta_1| + \sum_{i=2}^{k_2-1} l_i'' < l_{k_1}', \quad |\beta_1| + \sum_{i=2}^{k_2} l_i'' > l_{k_1}'.$$

Снова равенства быть не может, т. к. **в этом случае слово  $\beta$  можно было бы уменьшить**, что не так.

Тогда  $B_{k_1}' = \beta_1 B_{k_2-1}'' \dots B_2'' \beta_2$  для некоторого слова  $\beta_2 \in B^*$ ,  $\beta_2 \neq \Lambda$ .

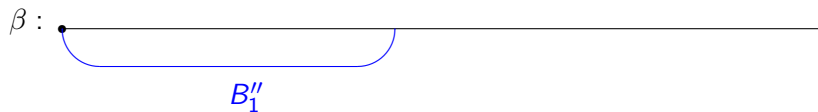
Слово  $\beta_2$  является собственным префиксом кодового слова  $B_{k_2}''$  и собственным суффиксом кодового слова  $B_{k_1}'$ , а также **не является никаким кодовым словом**.

Значит, в графе  $G_\varphi$  присутствует дуга  $e_2 = (\beta_1, \beta_2) \in E_\varphi$ , которой приписана пометка  $D_2 = B_{k_2-1}'' \dots B_2''$ .

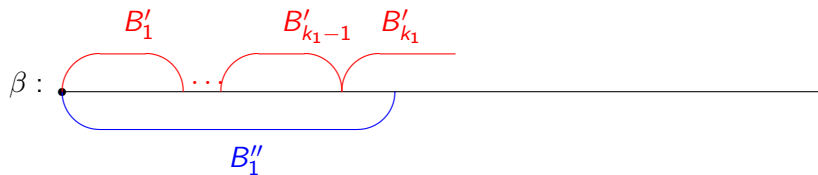
## Пояснение выбора числа $k_2$

$\beta$ : • \_\_\_\_\_

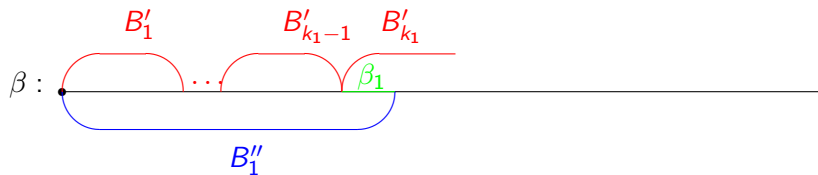
## Пояснение выбора числа $k_2$



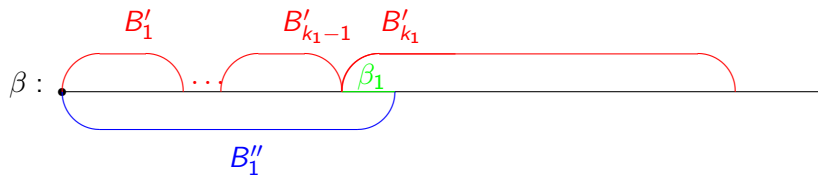
# Пояснение выбора числа $k_2$



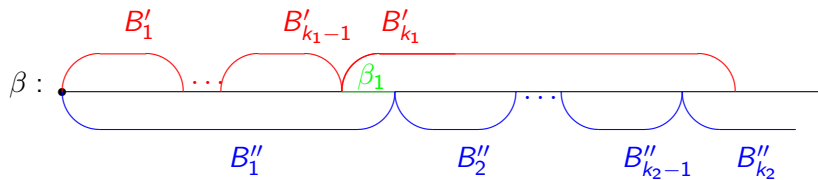
# Пояснение выбора числа $k_2$



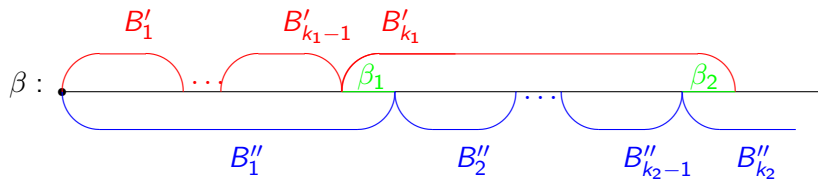
## Пояснение выбора числа $k_2$



# Пояснение выбора числа $k_2$



# Пояснение выбора числа $k_2$





# Критерий разделимости алфавитного кода

**Доказательство.** Далее найдем такое число  $k_3$ , что

$$|\beta_2| + \sum_{i=k_1+1}^{k_3-1} l'_i < l''_{k_2}, \quad |\beta_2| + \sum_{i=k_1+1}^{k_3} l'_i > l''_{k_2}.$$

Равенства быть не может, т. к. **в этом случае слово  $\beta$  можно было бы уменьшить**, что не так.

Тогда  $B''_{k_3} = \beta_2 B'_{k_1+1} \dots B'_{k_3-1} \beta_3$  для некоторого слова  $\beta_3 \in B^*$ ,  $\beta_3 \neq \Lambda$ .

Значит, в графе  $G_\varphi$  присутствует дуга  $e_3 = (\beta_2, \beta_3) \in E_\varphi$ , которой приписана пометка  $D_3 = B'_{k_1+1} \dots B'_{k_3-1}$ .

И т. д.

## Критерий разделимости алфавитного кода

**Доказательство.** Через конечное число таких шагов достигнем окончания слова  $\beta$ .

Значит, в графе  $G_\varphi$  присутствует дуга  $e_{m+1} = (\beta_m, \Lambda) \in E_\varphi$  для некоторого слова  $\beta_m \in B^*$ ,  $\beta_m \neq \Lambda$ .

Этой дуге  $e_{m+1}$  приписана пометка  $D_{m+1} = B_{k_{m-1}+1}^\circ \cdots B_{k_{m+1}-1}^\circ$ , где  $\circ \in \{', ''\}$  в зависимости от четности числа  $m$ .

Таким образом, в графе  $G_\varphi$  найдется ориентированный замкнутый путь:

$$P = \Lambda, e_1, \beta_1, e_2, \beta_2, \dots, \beta_m, e_{m+1}, \Lambda,$$

в котором вершина  $\Lambda$  не встречается среди вершин  $\beta_1, \dots, \beta_m$ .

Из этого пути  $P$  можно выделить **ориентированный цикл (в частности, петлю), проходящий через вершину  $\Lambda$ .**

## Критерий разделимости алфавитного кода

**Доказательство.** 2. Пусть теперь в графе  $G_\varphi$  найдется ориентированный цикл (в частности, петля)

$$P = \Lambda, e_1, \beta_1, e_2, \beta_2, \dots, \beta_m, e_{m+1}, \Lambda,$$

проходящий через вершину  $\Lambda$ .

Пусть дуге  $e_i$  приспана пометка  $D_i = B_{i_1}, \dots, B_{i_{k_i}}$ ,  
 $i = 1, \dots, m, m + 1$ .

Покажем, что слово

$$\beta = D_1\beta_1 D_2\beta_2 \dots \beta_m D_{m+1} \in B^*$$

допускает не менее двух декодирований.

# Критерий разделимости алфавитного кода

Доказательство. Итак, рассмотрим слово

$$\beta = D_1\beta_1 D_2\beta_2 \dots \beta_m D_{m+1} \in B^*.$$

Пусть, для определенности,  $m$  — четно.

Первое декодирование:

$$D_1\beta_1 D_2\beta_2 D_3\beta_3 D_4 \dots D_m\beta_m D_{m+1}.$$

Второе декодирование:

$$D_1\beta_1 D_2\beta_2 D_3\beta_3 D_4\beta_4 \dots \beta_{m-1} D_m\beta_m D_{m+1}.$$

Случай нечетного  $m$  разбирается аналогично.

Значит, код  $C_\varphi$  не является разделимым.



## Алгоритм проверки делимости алфавитного кода

*Вход:* алфавитный код  $C = \{B_1, \dots, B_r\} \subseteq B^*$  в кодирующем алфавите  $B$ .

*Выход:* «да», если код  $C$  является делимым, и «нет» и слово  $\beta \in B^*$ , допускающее не менее двух декодирований, в обратном случае.

## Проверка делимости алфавитного кода

*Описание алгоритма.*

1. Построить орграф  $G$  делимости кода  $C$ .
2. Если граф  $G$  не содержит петель или направленных циклов, проходящих через «пустую» вершину, то выдать «да» и остановиться.
3. Иначе, пусть  $\beta_0, \beta_1, \dots, \beta_m, \beta_0$  — направленный цикл в  $G$ , где  $\beta_i \in B^*$ ,  $i = 1, \dots, m$ ,  $\beta_0 = \Lambda$ , причем дуга  $(\beta_{i-1}, \beta_i)$  помечена последовательностью  $D_i$ ,  $i = 1, \dots, m$ , а дуга  $(\beta_m, \beta_0)$  помечена последовательностью  $D_{m+1}$ . Тогда выдать «нет» и

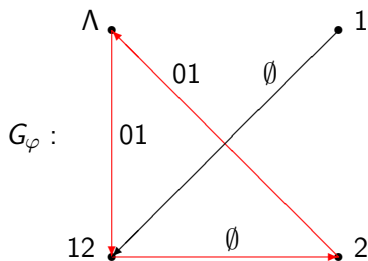
$$\beta = D_1\beta_1 D_2\beta_2 \dots \beta_m D_{m+1} \in B^*$$

и остановиться.

*Окончание описания алгоритма.*

# Проверка делимости алфавитного кода

**Пример.** Рассмотрим код  $C_\varphi = \{01, 201, 112, 122, 0112\}$ .



Получаем:

$$\beta = 0112201 = 0112 + 201 = 01 + 122 + 01.$$

1. Алексеев В.Б. Лекции по дискретной математике. М.: Инфра-М, 2012. С. 41–46.