Лекция 12. Кодирование. Алфавитные коды. Равномерные и префиксные коды. Алгоритм проверки разделимости алфавитного кода. Теорема Маркова.

Лектор — Селезнева Светлана Николаевна
selezn@cs.msu.ru

факультет ВМК МГУ имени М.В. Ломоносова

Алфавит

Алфавитом называется произвольное конечное множество.

Если A — алфавит, то любой элемент $a \in A$ называется **буквой** алфавита A.

Например, $A = \{0, 1\}$ — алфавит из двух букв: 0 и 1.

Конечные слова в алфавите

Пусть A — алфавит.

(**Конечным**) **словом** в алфавите A назовем произвольную конечную последовательность букв из A.

Введем обозначение: Λ — **пустое** слово (т. е. слово без букв).

Множество всех конечных слов в алфавите A обозначим A^*.

Длиной $|\alpha|$ слова $\alpha \in A^*$ назовем число букв в нем, $|\Lambda| = 0$.

Если $\alpha \in A^*$, то i-ю букву слова α обозначим $\alpha(i)$, $i = 1, \ldots, |\alpha|$.

Т. е. если $|\alpha| = m$, где $m \geqslant 1$, то $\alpha = \alpha(1)\alpha(2)\ldots\alpha(m)$.

Соединение слов

Если A — алфавит и $\alpha, \beta \in A^*$, где $|\alpha| = k$, $|\beta| = m$, то **соединением** (или **конкатенацией**) слов α и β назовем слово $\alpha\beta \in A^*$, где

$$\alpha\beta = \alpha(1)\ldots\alpha(k)\beta(1)\ldots\beta(m).$$

Отметим, что $|\alpha\beta| = k + m$.

Подслово

Пусть A — алфавит. Слово $\beta \in A^*$ называется **подсловом** слова $\alpha \in A^*$, если найдутся такие слова $\alpha_1, \alpha_2 \in A^*$ (возможно, пустые), что

$$\alpha = \alpha_1 \beta \alpha_2.$$

Префиксы слова

Пусть A — алфавит и $\alpha \in A^*$, $\alpha = \alpha(1)\alpha(2)\dots\alpha(m)$.

Префиксом слова α называется любое его начало, в том числе, и пустое. Т. е. если $\alpha = \alpha_1\alpha_2$, где $\alpha_1, \alpha_2 \in A^*$, то α_1 — префикс слова α.

Значит, всеми возможными префиксами слова α являются слова:

$$\Lambda,\ \alpha(1),\ \alpha(1)\alpha(2),\ \dots,\ \alpha(1)\alpha(2)\dots\alpha(m).$$

Например, если $A = \{0, 1\}$ и $\alpha = 0011$, то

$$\Lambda,\ 0,\ 00,\ 001,\ 0011\ -$$

все префиксы слова α.

Префикс называется **собственным**, если он не совпадает ни с пустым словом, ни со словом α.

Суффиксы слова

Пусть A — алфавит и $\alpha \in A^*$, $\alpha = \alpha(1)\alpha(2)\dots\alpha(m)$.

Суффиксом слова α называется любое его окончание, в том числе, и пустое. Т. е. если $\alpha = \alpha_1\alpha_2$, где $\alpha_1, \alpha_2 \in A^*$, то α_2 — суффикс слова α.

Значит, всеми возможными суффиксами слова α являются слова:

$$\Lambda,\ \alpha(m),\ \alpha(m-1)\alpha(m),\ \dots,\ \alpha(1)\alpha(2)\dots\alpha(m).$$

Например, если $A = \{0, 1\}$ и $\alpha = 0011$, то

$$\Lambda,\ 1,\ 11,\ 011,\ 0011 -$$

все префиксы слова α.

Суффикс называется **собственным**, если он не совпадает ни с пустым словом, ни со словом α.

Кодирование

Пусть заданы два алфавита A и B.

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

Кодированием (из A в B) называется произвольное отображение

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

При кодировании φ любое слово $\alpha \in A^*$ называется **сообщением**, а слово $\beta = \varphi(\alpha) \in B^*$ — его **кодом**.

Можно рассматривать кодирования вида $\varphi : S \to B^*$, где $S \subseteq A^*$ — множество (исходных) сообщений.

Код

Если $\varphi : A^* \to B^*$ — кодирование, то множество кодов всех слов из A^* назовем **кодом** C_φ, т. е.

$$C_\varphi = \{\varphi(\alpha) \mid \alpha \in A^*\} \subseteq B^*.$$

Т. е. код C_φ — множество кодов всех сообщений.

Кодирования

Как правило (в приложениях), **кодирование** $\varphi : A^* \to B^*$ описывается **схемой (или алгоритмом) кодирования**, т. е. правилом, позволяющим для любого сообщения $\alpha \in A^*$ получить его код $\varphi(\alpha) \in B^*$.

Кроме того, требуется **схема (или алгоритм) декодирования**, т. е. правило, позволяющее для любого слова $\beta \in B^*$ понять, является ли β кодом какого-то сообщения, и при положительном ответе найти такое сообщение $\alpha \in A^*$, что $\varphi(\alpha) = \beta$.

Очень желательным свойством для кодирования φ является однозначность декодирования.

Разделимость кодирования

Кодирование $\varphi : A^* \to B^*$ 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 φ — разделимо, если оно разным сообщениям сопоставляет различные коды.

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

Алфавитное кодирование

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

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

1) заданы различные непустые коды букв алфавита A:

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

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

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

Алфавитный код

Пусть φ — алфавитное кодирование из A в B, т. е.

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

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

Множество всех кодовых слов при кодировании φ назовем **алфавитным кодом** C_φ, т. е.

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

Отметим, что код C_φ однозначно определяет алфавитное кодирование φ (при заданном порядке букв из A).

Алфавитный код C_φ назовем **однозначным** (или **разделимым**), если кодирование φ — разделимо.

Декодирование

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

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

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

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

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

Алфавитные коды

Пример. Пусть $A = \{a_1, a_2, a_3\}$, $B = \{0, 1\}$ и φ_1 — алфавитное кодирование из A в B, где

$$\begin{aligned} \varphi_1(a_1) &= 00, \\ \varphi_2(a_2) &= 01, \\ \varphi_3(a_3) &= 11. \end{aligned}$$

Тогда, например, для сообщения $\alpha = a_1 a_1 a_3 \in A^*$ его код:

$$\varphi_1(\alpha) = 000011 \in B^*.$$

Код C_{φ_1} является разделимым (почему?).

Равномерный алфавитный код

Алфавитный код $C = \{B_1, \ldots, B_r\} \subseteq B^*$ называется **равномерным**, если длины всех его кодовых слов одинаковы, т. е.

$$|B_1| = |B_2| = \ldots = |B_r|.$$

Разделимость равномерного алфавитного кода

Предложение 12.1. *Любой равномерный алфавитный код является разделимым*.

Доказательство. Пусть $C = \{B_1, \ldots, B_r\} \subseteq B^*$ — равномерный алфавитный код, $|B_i| = l$, $i = 1, \ldots, r$.

Рассмотрим произвольное слово $\beta \in B^*$. Как можно его декодировать?

Разобьем слово β на последовательность слов длины l.

Если такое разбиение возможно, и каждое слово в этой последовательности является кодовым, то получили однозначное декодирование слова β.

Иначе, слово β не допускает декодирования.

□

Алфавитные коды

Пример. Пусть $A = \{a_1, a_2, a_3\}$, $B = \{0, 1\}$ и φ_2 — алфавитное кодирование, где

$$\begin{aligned} \varphi_2(a_1) &= 0, \\ \varphi_2(a_2) &= 10, \\ \varphi_2(a_3) &= 112. \end{aligned}$$

Тогда, например, для сообщения $\alpha = a_1 a_1 a_3 \in A^*$ его код:

$$\varphi_2(\alpha) = 00110 \in B^*.$$

Код C_{φ_2} является разделимым (почему?).

Префиксный алфавитный код

Алфавитный код $C = \{B_1, \ldots, B_r\} \subseteq B^*$ называется **префиксным**, если никакое его кодовое слово не является префиксом никакого другого его кодового слова, т. е.

$$\nexists\, B_i, B_j \in C : B_i = B_j\beta_2 \text{ для некоторого слова } \beta_2 \in B^*.$$

Разделимость префиксного алфавитного кода

Предложение 12.2. *Любой префиксный алфавитный код является разделимым*.

Доказательство. Пусть $C = \{B_1, \ldots, B_r\} \subseteq B^*$ — префиксный алфавитный код.

Рассмотрим произвольное слово $\beta \in B^*$. Как можно его декодировать?

Разделимость префиксного алфавитного кода

Доказательство. Пусть $\beta = B_{i_1}\beta_1$ для некоторого кодового слова $B_{i_1} \in C$ и некоторого слова $\beta_1 \in B^*$.

Отметим, что кодовое слово B_{i_1} (если оно существует) находится однозначно.

Действительно, если

$$\beta = B_{i_1}\beta_1 = B_{j_1}\beta_1',$$

где $B_{i_1}, B_{j_1} \in C$, $B_{i_1} \neq B_{j_1}$, $\beta_1, \beta_1' \in B^*$, то какое-то из кодовых слов B_{i_1}, B_{j_1} является префиксом другого, что невозможно.

Далее повторим рассуждения для слова $\beta_1 \in B^*$.

В итоге, либо однозначно декодируем слово β, либо на каком-то шаге не найдем подходящее кодовое слово, а значит, слово β не допускает декодирования.

Суффиксный алфавитный код

Алфавитный код $C = \{B_1, \ldots, B_r\} \subseteq B^*$ называется **суффиксным**, если никакое его кодовое слово не является суффиксом никакого другого его кодового слова, т. е.

$$\nexists\, B_i, B_j \in C : B_i = \beta_1 B_j \text{ для некоторого слова } \beta_1 \in B^*.$$

Предложение 12.3. *Любой суффиксный алфавитный код является разделимым*.

Доказательство проводится подобно доказательству предыдущего утверждения, только начинать декодировать слово β нужно справа.

Алфавитные коды

Пример. Пусть $A = \{a_1, a_2, a_3\}$, $B = \{0, 1\}$ и φ_3 — алфавитное кодирование, где

$$\begin{aligned} \varphi_3(a_1) &= 0, \\ \varphi_3(a_2) &= 1, \\ \varphi_3(a_3) &= 01. \end{aligned}$$

Тогда, например, для сообщения $\alpha = a_1 a_1 a_3 \in A^*$ его код:

$$\varphi_3(\alpha) = 0001 \in B^*.$$

Код C_{φ_3} не является разделимым, т. к. слово $\beta = 01 \in B^*$ можно декодировать двумя способами:

$$\begin{aligned} \beta &= 01, \\ \beta &= 01, \end{aligned}$$

т. е. $\beta = \varphi_3(a_1 a_2) = \varphi_3(a_3)$.

Граф однозначности алфавитного кода

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

Построим *орграф* $G_\varphi = (V_\varphi, E_\varphi)$ для кода C_φ.

1. Множество вершин V_φ, $V_\varphi \subseteq B^*$, состоит из пустого слова Λ 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.

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

Граф однозначности алфавитного кода

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

2. Опишем множество дуг E_φ: 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 $B_{i_1}, \ldots, B_{i_k}$, что

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

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

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

Граф однозначности алфавитного кода

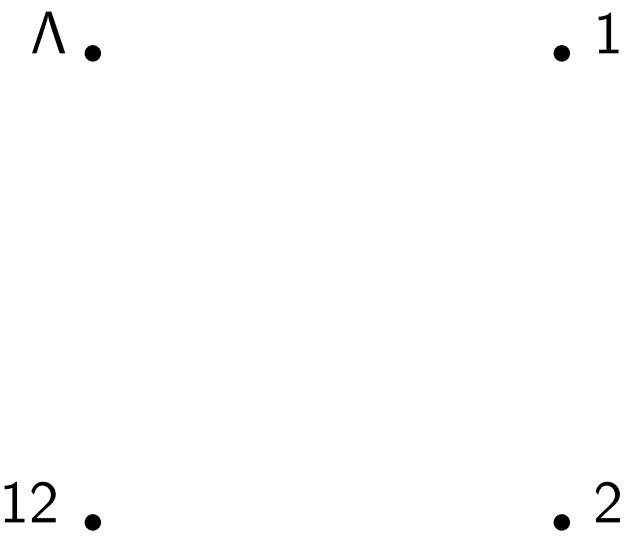
Пример. Пусть $C_\varphi = \{01, 201, 112, 122, 0112\}$. Построим граф $G_\varphi = (V_\varphi, E_\varphi)$.

Граф однозначности алфавитного кода

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

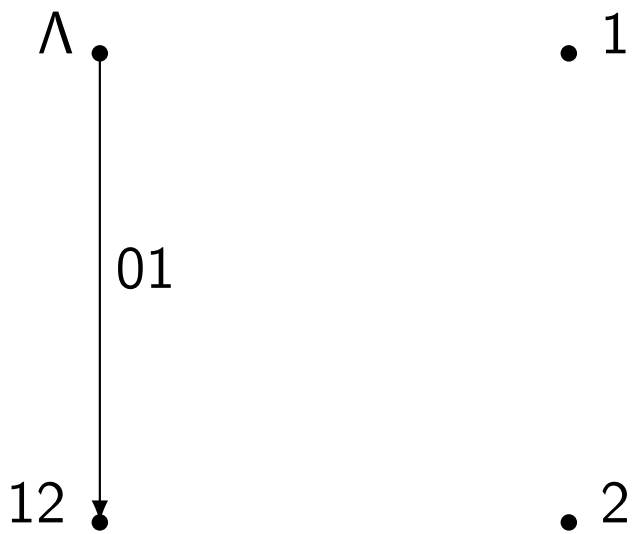
Граф однозначности алфавитного кода

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



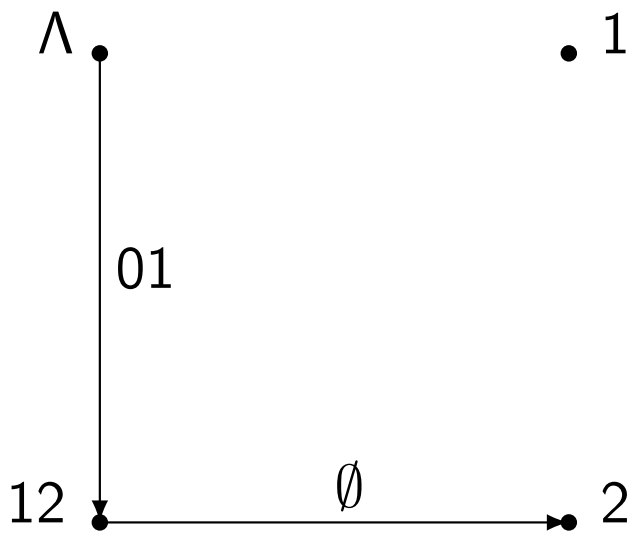
Граф однозначности алфавитного кода

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



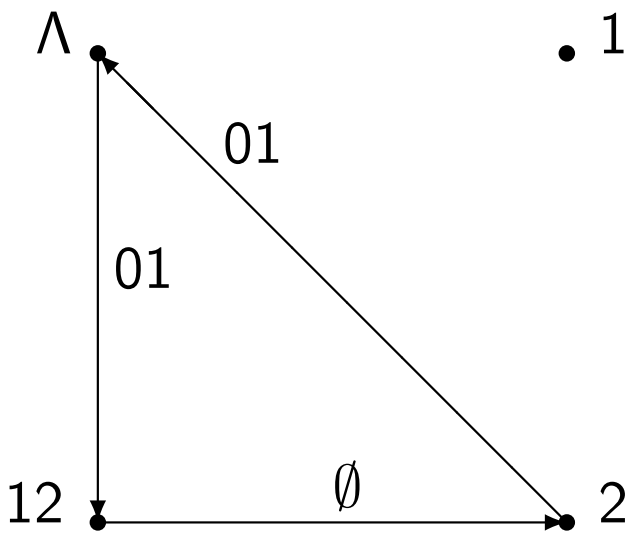
Граф однозначности алфавитного кода

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



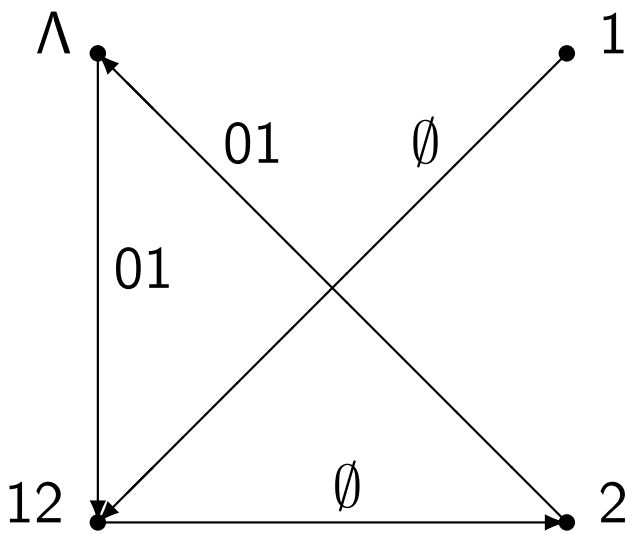
Граф однозначности алфавитного кода

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



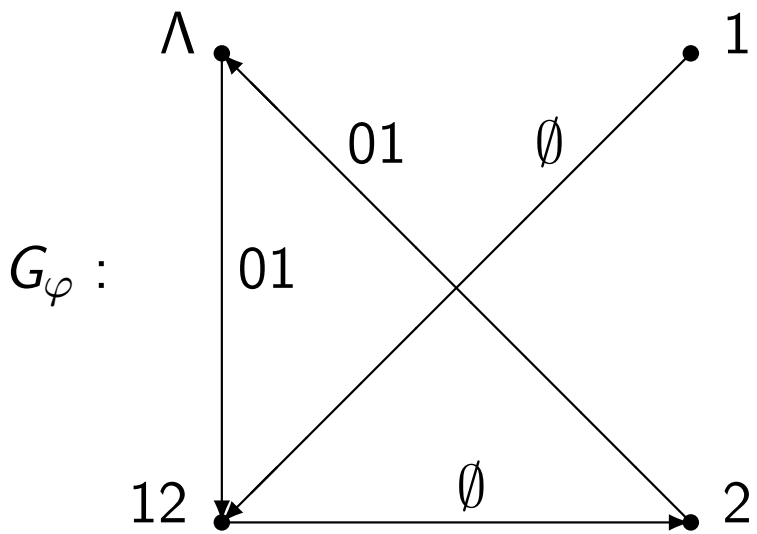
Граф однозначности алфавитного кода

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



Граф однозначности 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 12.1. *Алфавитный код C_φ является разделимым тогда и только тогда, когда в графе G_φ отсутствуют ориентированные циклы (в том числе, и петли), проходящие через вершину Λ.*

Доказательство. Пусть $C_\varphi = \{B_1, \ldots, B_r\} \subseteq B^*$ — алфавитный код и G_φ — граф для кода C_φ.

Критерий разделимости алфавитного кода

Доказательство. 1. Пусть код C_φ 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 β на кодовые слова в 1-м декодировании и $\beta = B''_1 B''_2 \dots B''_{t_2}$ — разбиение слова β 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 разделимости алфавитного кода

Доказательство. Найдем такое число k_1, что

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

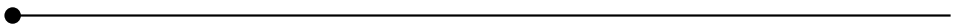
Заметим, что равенства здесь быть не может, т. к. в этом случае слово β можно было бы уменьшить, что не так.

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

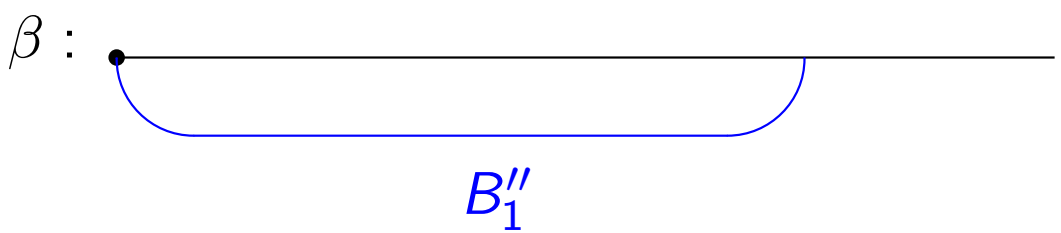
Отметим, что слово β_1 является собственным префиксом кодового слова B'_{k_1} и собственным суффиксом кодового слова B''_1, а также не является никаким кодовым словом.

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

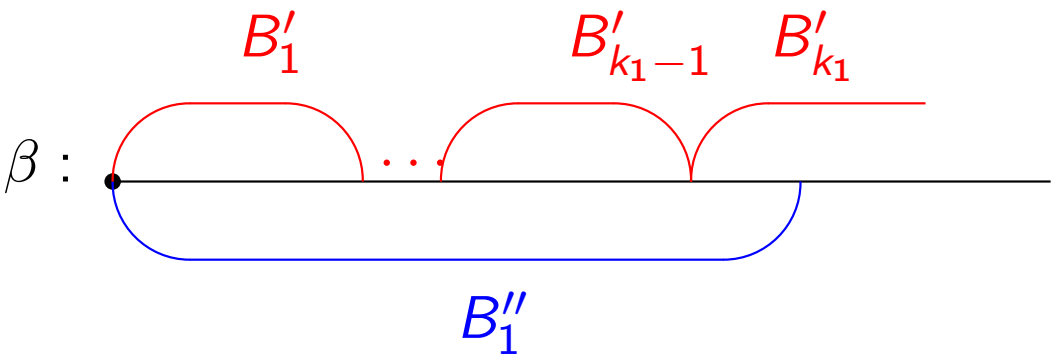
Пояснение выбора числа k_1

β : 

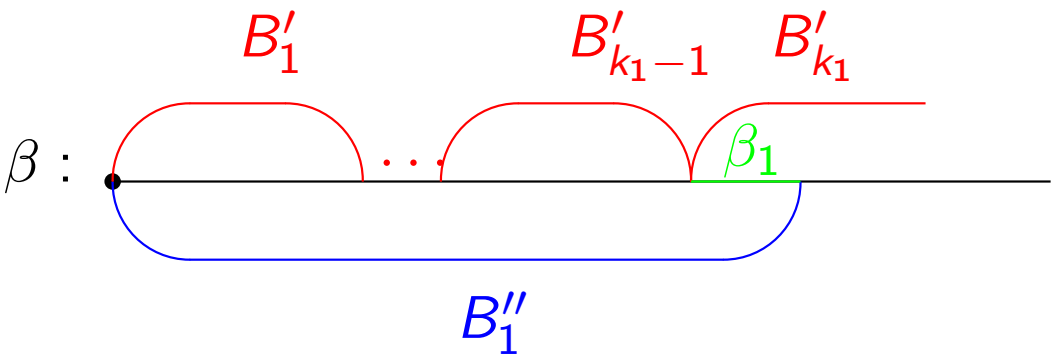
Пояснение выбора числа k_1



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 β можно было бы уменьшить, что не так.

Тогда $B'_{k_1} = \beta_1 B''_2 \dots B''_{k_2-1}\beta_2$ для некоторого слова $\beta_2 \in B^*$, $\beta_2 \neq \Lambda$.

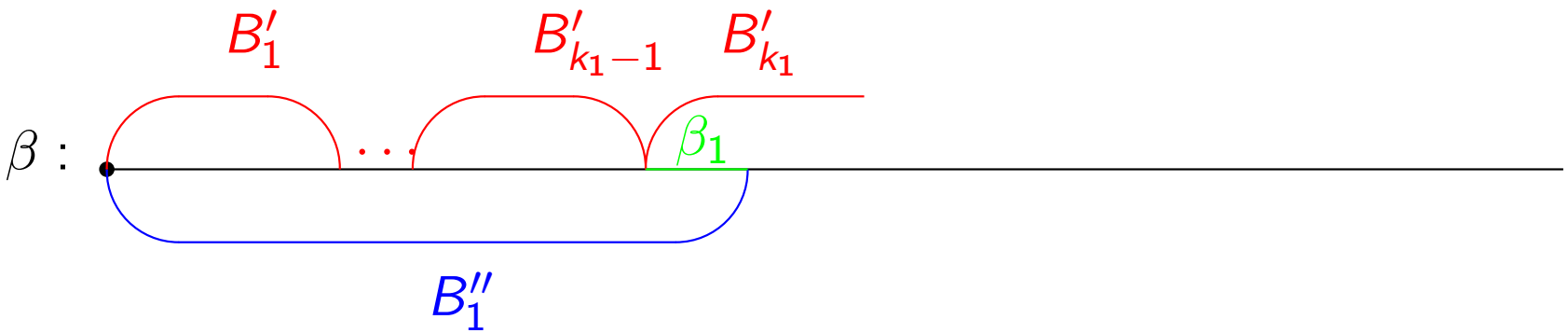
Слово β_2 является собственным префиксом кодового слова B''_{k_2} и собственным суффиксом кодового слова B'_{k_1}, а также не является никаким кодовым словом.

Значит, в графе G_φ присутствует дуга $e_2 = (\beta_1, \beta_2) \in E_\varphi$, которой приписана пометка $D_2 = B''_2 \dots B''_{k_2-1}$.

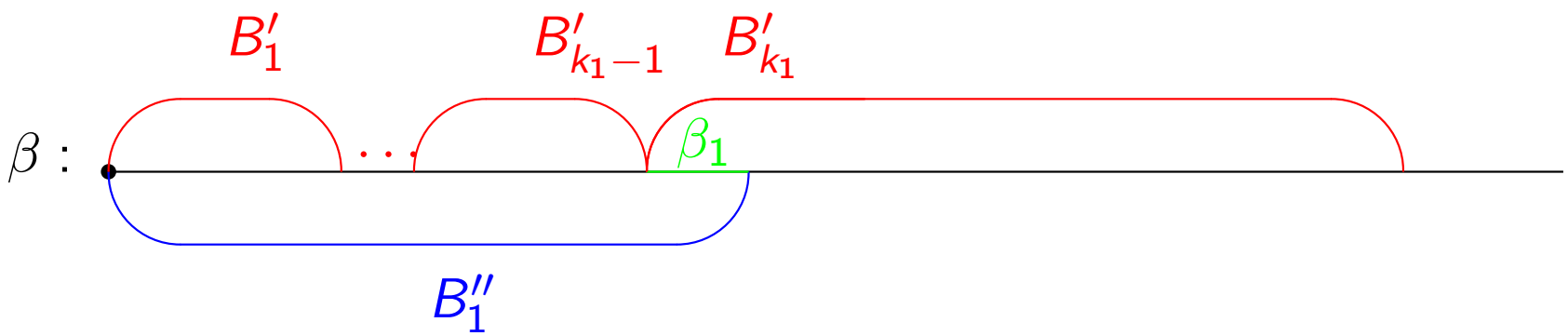
Пояснение выбора числа k_2

β : 

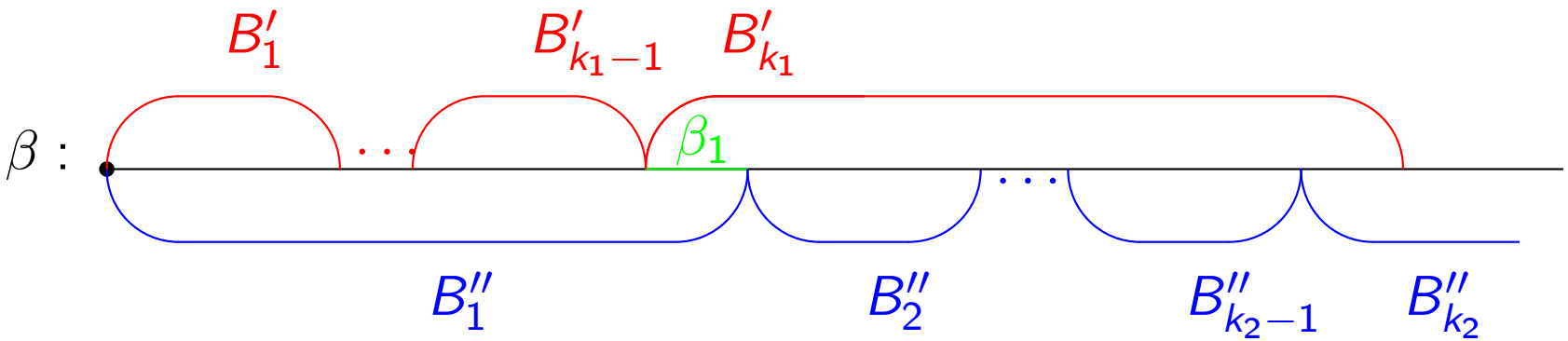
Пояснение выбора числа k_2



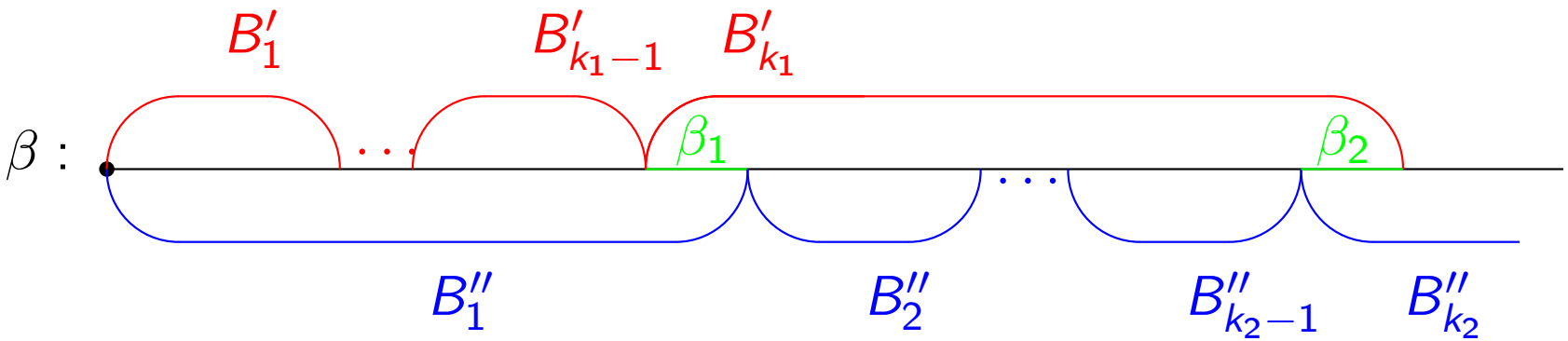
Пояснение выбора числа k_2



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 β 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 $\beta_3 \in B^*$, $\beta_3 \neq \Lambda$.

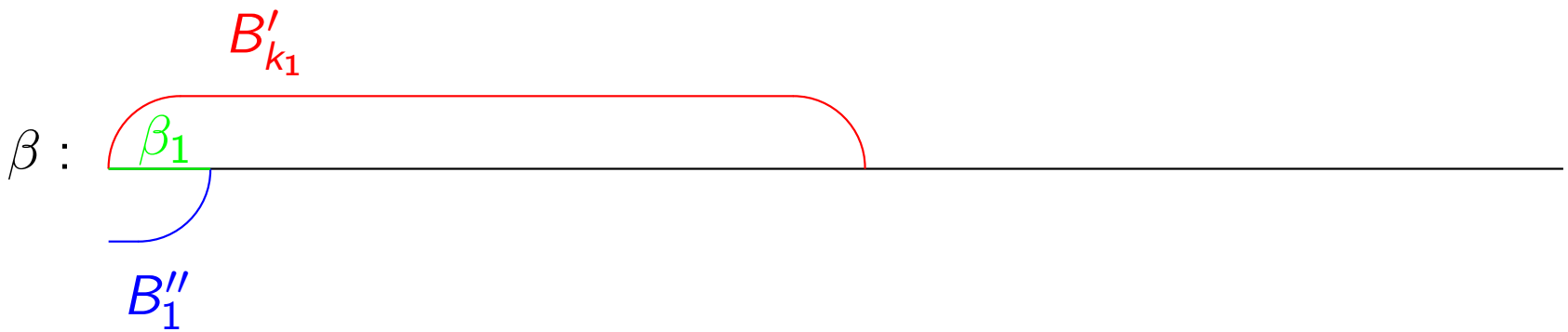
Значит, в графе G_φ присутствует дуга $e_3 = (\beta_2, \beta_3) \in E_\varphi$, которой приписана пометка $D_3 = B'_{k_1+1} \dots B'_{k_3-1}$.

И т. д.

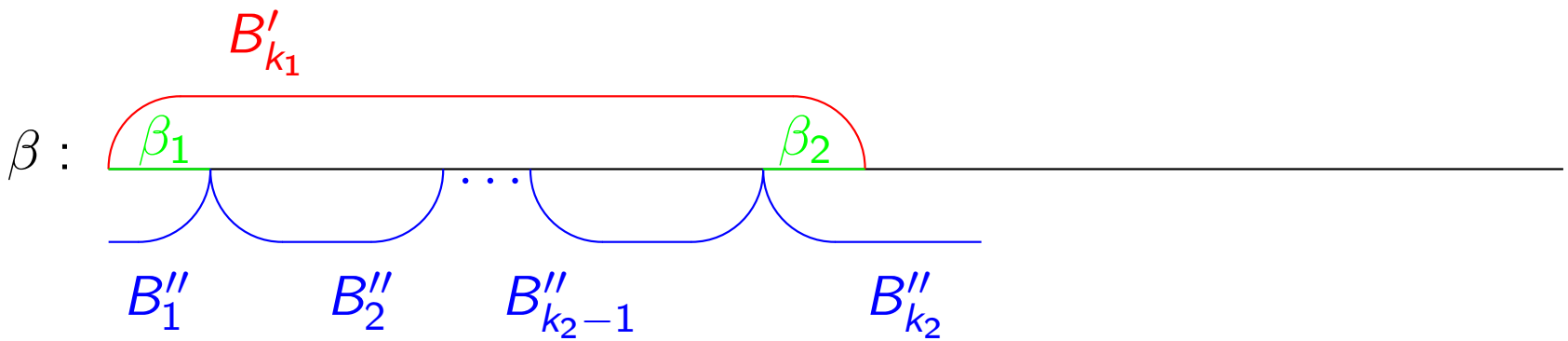
Пояснение выбора числа k_3

β : _____

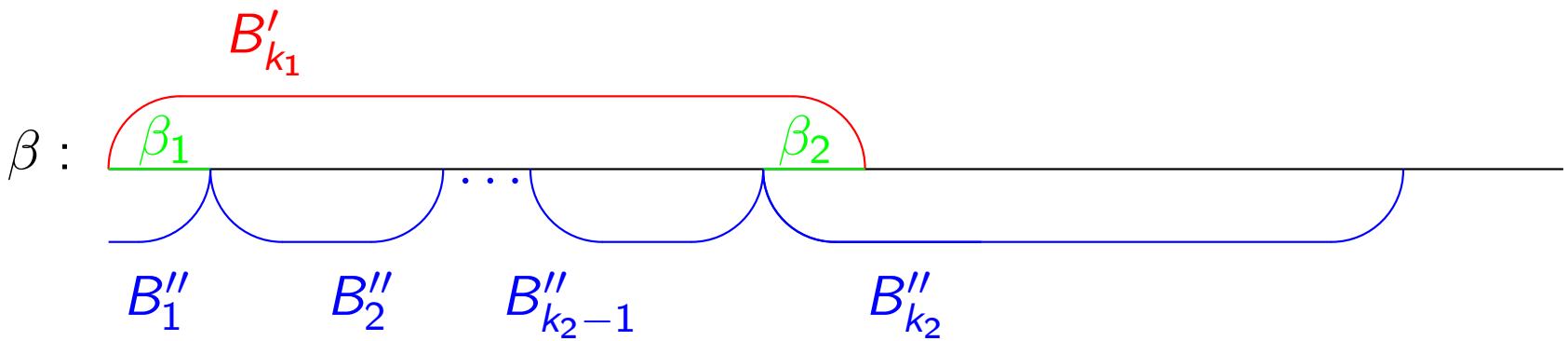
Пояснение выбора числа k_3



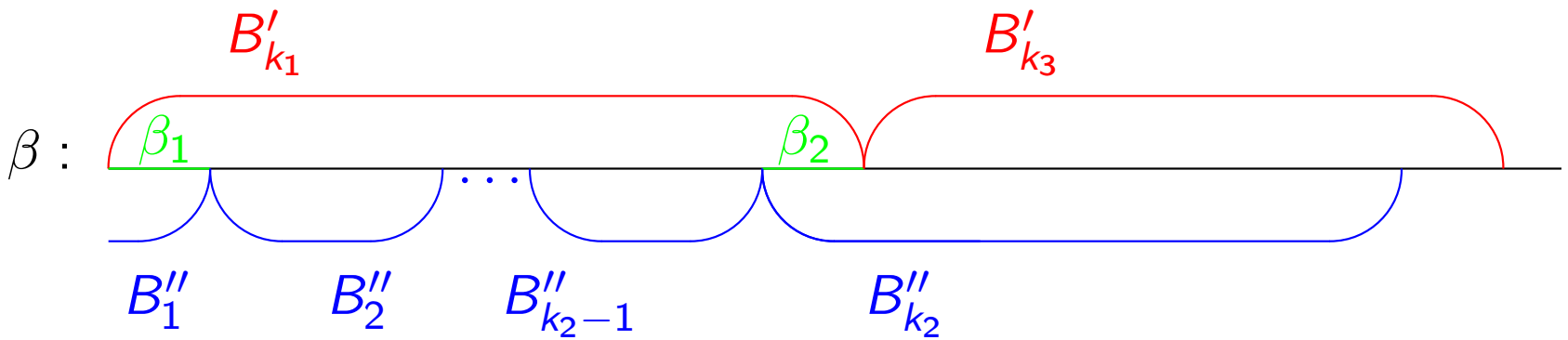
Пояснение выбора числа k_3



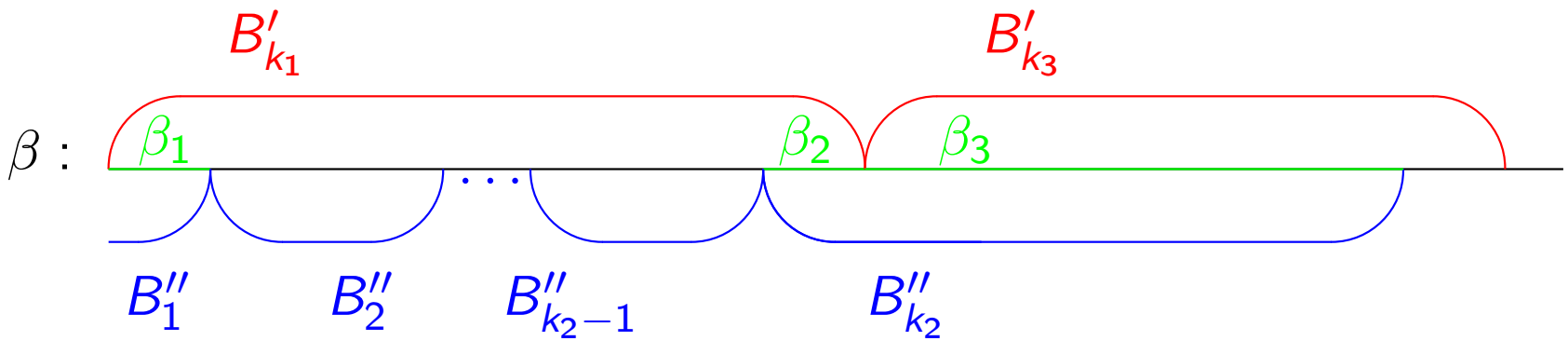
Пояснение выбора числа k_3



Пояснение выбора числа k_3



Пояснение выбора числа k_3



Критерий разделимости алфавитного кода

Доказательство. Через конечное число число таких шагов достигнем окончания слова β.

Значит, в графе G_φ 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 $D_{m+1} = B^\circ_{k_{m-1}+1} \dots B^\circ_{k_{m+1}-1}$, где $\circ \in \{', ''\}$ в зависимости от четности числа m.

Таким образом, в графе G_φ найдется ориентированный замкнутый путь:

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

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

Из этого пути P можно выделить ориентированный цикл (в частности, петлю), проходящий через вершину Λ.

Критерий разделимости алфавитного кода

Доказательство. 2. Пусть теперь в графе G_φ найдется ориентированный цикл (в частности, петля)

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

проходящий через вершину Λ.

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

Покажем, что слово

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

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_φ не является разделимым.

□

Проверка разделимости алфавитного кода

Алгоритм проверки разделимости алфавитного кода

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

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

Проверка разделимости алфавитного кода

Описание алгоритма.

1. Построить орграф G для кода C.

2. Если граф G не содержит петель или направленных циклов, проходящих через «пустую» вершину, то выдать «да» и остановиться.

3. Иначе, пусть $\beta_0, \beta_1, \ldots, \beta_m, \beta_0$ — направленный цикл в G, где $\beta_i \in B^*$, $i = 1, \ldots, m$, $\beta_0 = \Lambda$, причем дуга (β_{i-1}, β_i) помечена последовательностью D_i, $i = 1, \ldots, m$, а дуга (β_m, β_0) помечена последовательностью D_{m+1}. Тогда выдать «нет» и

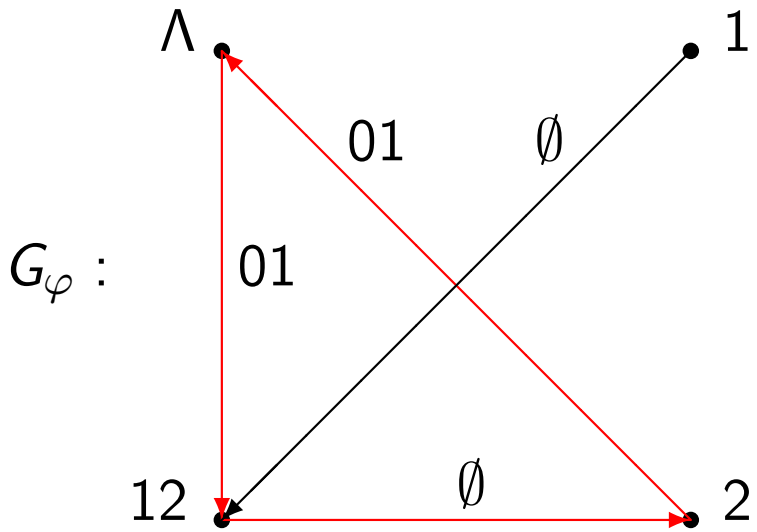
$$\beta = D_1\beta_1 D_2\beta_2 \ldots \beta_m D_{m+1} \in B^*$$

и остановиться.

Окончание описания алгоритма.

Проверка разделимости алфавитного кода

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



Получаем:

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

Теорема Маркова

Теорема 12.2 (А. А. Марков). *Пусть A — исходный алфавит, $C_\varphi = \{B_1, \ldots, B_r\} \subseteq B^*$ — алфавитный код, где $|B_i| = l_i$, $i = 1, \ldots, r$. Пусть $L = \sum_{i=1}^{r} l_i$ и w обозначает наибольшее число кодовых слов (возможно, с повторами), соединение которых является подсловом какого-то кодового слова. Тогда если код C_φ не является разделимым, то найдутся такие слова $\alpha_1, \alpha_2 \in A^*$, $\alpha_1 \neq \alpha_2$, $\varphi(\alpha_1) = \varphi(\alpha_2)$, что*

$$|\alpha_1|, |\alpha_2| \leqslant \lfloor \frac{(L - r + 2)(w + 1)}{2} \rfloor,$$

где $\lfloor a \rfloor$ обозначает целую часть числа a.

Теорема Маркова

Доказательство. Код C_φ не является разделимым, значит, в графе G_φ найдется ориентированный цикл (в частности, петля)

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

проходящий через вершину Λ.

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

Можно считать, что P — петля или простой цикл, поэтому слова $\beta_1, \ldots, \beta_m$ — различны.

Каждое слово β_i является, в частности, собственным префиксом какого-то кодового слова, поэтому

$$m \leqslant \sum_{i=1}^{r} (l_i - 1) = L - r.$$

Теорема Маркова

Доказательство. Рассмотрим слово

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

которое допускает не менее двух декодирований. Пусть $\alpha_1, \alpha_2 \in A^*$, $\alpha_1 \neq \alpha_2$, — два декодирования слова β, т. е.

$$\beta = \varphi(\alpha_1) = \varphi(\alpha_2).$$

Слова $\beta_1, \dots, \beta_m$ разбивают слово β на $m+1$ частей: $D_1, \dots, D_{m+1}$.

Рассмотрим k пар частей:

$$(D_1, D_2),\ (D_3, D_4),\ \dots,\ (D_{2k-1}, D_{2k}),$$

где $k = \lfloor \frac{m+1}{2} \rfloor$.

Теорема Маркова

Доказательство. Для каждого $i = 1, \ldots, k$ слова

$$\begin{aligned} \beta'_i &= \overbrace{\beta_{2i-2} D_{2i-1} \beta_{2i-1}}^{1} \overbrace{D_{2i}}^{\leqslant w}, \\ \beta''_i &= \underbrace{D_{2i-1}}_{\leqslant w} \underbrace{\beta_{2i-1} D_{2i} \beta_{2i}}_{1}, \end{aligned}$$

разбиваются не более, чем на $w + 1$, кодовых слов.

Значит, каждая пара (D_{2i-1}, D_{2i}) вносит не более $w + 1$ кодовых слов в каждое из декодирований слова β.

Если $m + 1$ — нечетно, то останется еще последовательность D_{m+1}, которая также вносит не более $w + 1$ кодовых слов в каждое из декодирований слова β.

Теорема Маркова

Доказательство. Значит,

$$|\alpha_1|, |\alpha_2| \leqslant \frac{m+2}{2} \cdot (w+1) \leqslant \frac{(L-r+2)(w+1)}{2}.$$

Из того, что $|\alpha_1|, |\alpha_2|$ — целые числа, получаем утверждение теоремы.

□

Задачи для самостоятельного решения

1. Докажите предложение 12.3.

2. Разберите случай нечетного числа m в доказательстве теоремы 12.1.

3. Объясните, почему можно считать, что P — петля или простой цикл, в доказательстве теоремы 12.2.