Задание 1. Действия над словами

Текст на любом языке: языке программирования, естественном (национальном языке), языке математики или языке математической логики представляет собою цепочку символов (знаков). В виде цепочек, состоящих из символов "0" и "1" выражается и вся информация, циркулирующая в компьютере. Математика, посвящённая изучению цепочек символов — символьная математика — играет важную роль в понимании информационных процессов.

Учебные задачи

1. Овладеть терминами: *буква, алфавит, символьная фигура, слово в алфавите, пустое слово, однобуквенное слово, длина слова, подслово, вхождение слова, начало слова, конец слова, операция приписывания слова к слову*.

2. Научиться выполнять: *замену вхождения подслова, подстановку слова вместо буквы одновременную подстановку слов вместо букв*.

ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Исходные допущения. Слова "знак" "буква", "символ" считаются синонимами. Понятие *знака* считается известным заранее; оно является абстрактным образом тех материальных знаков, которые мы изображаем, например, на листе бумаги. Предполагается, что мы можем различать знаки и их *экземпляры*, записывать и удалять их. Предполагается, что знаки можно записывать упорядоченно в пределах некоторой идеальной (т.е.мыслимой нами) плоскости; мы будем использовать как линейный способ записи слева направо, так и (для удобства записи) упорядоченное расположение знаков на этой плоскости сверху вниз. Любую запись на плоскости, состоящую из символов мы будем называть (вслед за Герхардом Генценом [1967, с.11]) *фигурой*.

Понятие слова в алфавите. Произвольное непустое множество знаков называется *алфавитом*. Знаки, принадлежащие некоторому алфавиту, называют также *буквами* этого алфавита. Произвольная конечная последовательность, состоящая из букв алфавита, называется *словом*[1] *в* данном *алфавите* (или просто "словом"). Слово записывают слева направо без каких-либо разделительных знаков в виде $a_1a_2\ldots a_n$ (a_1, a_2, ..., a_n — буквы алфавита, не обязательно различные). Два слова $a_1\ldots a_n$ и $b_1\ldots b_k$ в произвольном алфавите A считаются *равными*, или, более точно, *графически равными*, если $n = k$ и $a_1 = b_1$, ..., $a_n = b_n$. В этом случае будем писать: $a_1\ldots a_n \doteq b_1\ldots b_k$. Число n называется *длиной слова* $a_1a_2\ldots a_n$. Длину произвольного слова $\boldsymbol{\alpha}$ будем обозначать $\mathrm{len}(\boldsymbol{\alpha})$. Допускается слово, не содержащее ни одного знака, его называют *пустым словом*. Длина пустого слова считается равной 0. Пустое слово обозначается Λ.

[1] Удобным термином, помогающим создать адекватный образ, может служить термин "*цепочка*". Программисты для используют для рассматриваемого понятия термин "*строка*" (*строка символов*).

Пусть $\boldsymbol{A}$ — произвольный алфавит. Множество всех слов в алфавите $\boldsymbol{A}$ (включая и пустое слово) будем обозначать $\boldsymbol{A}^\star$, а множество всех непустых слов в этом же алфавите — $\boldsymbol{A}^+$.

Операция приписывания слов. Основная операция в множестве $\boldsymbol{A}^\star$ — операция *приписывания* слова к слову: если дано слово $\boldsymbol{\alpha}$, имеющее вид $a_0\ldots a_n$ и слово $\boldsymbol{\beta}$ вида $b_0\ldots b_m$, то записав непосредственно после слова $\boldsymbol{\alpha}$ слово $\boldsymbol{\beta}$, получим новое слово

$$\boldsymbol{\alpha\beta} \doteq a_0\ldots a_n b_0\ldots b_m\,.$$

Слово $\boldsymbol{\alpha\beta}$ называют *результатом приписывания к слову* $\boldsymbol{\alpha}$ *слова* $\boldsymbol{\beta}$. В дальнейшем будем использовать без всяких ссылок свойства *ассоциативности* и *сократимости* операции приписывания слов.

Свойство ассоциативности приписывания слов состоит в следующем. Если к слову $\boldsymbol{\alpha}$ приписать слово $\boldsymbol{\beta}$, а затем к полученному в результате такого приписывания слову приписать некоторое слово γ, то получится тот же самый результат, если к слову $\boldsymbol{\alpha}$ приписать результат приписывания к слову $\boldsymbol{\beta}$ слова γ.

Свойство сократимости операции ассоциативности приписывания слов состоит в следующем. Если для некоторых слов $\boldsymbol{\alpha}, \boldsymbol{\beta}$ и $\boldsymbol{\gamma}$ выполняется $\boldsymbol{\alpha\gamma} \doteq \boldsymbol{\beta\gamma}$, то $\boldsymbol{\alpha} \doteq \boldsymbol{\beta}$ (сократимость справа). Аналогично и с другой стороны: если $\boldsymbol{\gamma\alpha} \doteq \boldsymbol{\gamma\beta}$, то $\boldsymbol{\alpha} \doteq \boldsymbol{\beta}$ (сократимость слева).

Подслова и их вхождения в слово. Подслово — это часть слово. Математически строго подслово можно определить так. Слово $\boldsymbol{\beta}$ называется *подсловом слова* $\boldsymbol{\alpha}$ ($\boldsymbol{\alpha}, \boldsymbol{\beta} \in \boldsymbol{A}^\star$), если найдутся такие слова $\boldsymbol{\gamma}, \boldsymbol{\delta} \in \boldsymbol{A}^\star$, что $\boldsymbol{\alpha} \doteq \boldsymbol{\gamma\beta\delta}$.

Пусть для некоторых слов выполняется $\boldsymbol{\alpha} \doteq \boldsymbol{\gamma\beta\delta} \doteq \boldsymbol{\gamma_1\beta\delta_1}$ и при этом $\boldsymbol{\gamma}$ графически не равно слову $\boldsymbol{\gamma_1}$: $\boldsymbol{\gamma} \not\doteq \boldsymbol{\gamma_1}$. В этом случае говорят, что мы имеем два *различных вхождения* подслова $\boldsymbol{\beta}$ в слово $\boldsymbol{\alpha}$. Таким образом, *вхождением слова* $\boldsymbol{\beta}$ *в слово* $\boldsymbol{\alpha}$ называют подслово $\boldsymbol{\beta}$ вместе с его *местом* в слове $\boldsymbol{\alpha}$. Более точно, *вхождение слова* $\boldsymbol{\beta}$ *в слово* $\boldsymbol{\alpha}$ можно определить как тройку слов $(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\delta})$ такую, что $\boldsymbol{\alpha} \doteq \boldsymbol{\gamma\beta\delta}$.

Вхождением буквы a *в слово* называется вхождение в это слово слова, состоящего из одной буквы a (т.е. *однобуквенного слова*). Если существует вхождение буквы a в слово $\boldsymbol{\alpha}$, то говорим, что *буква a входит в* $\boldsymbol{\alpha}$.

Если $\boldsymbol{\alpha} \doteq \boldsymbol{\gamma_0\beta\delta_0}$, и при этом слово $\boldsymbol{\gamma_0}$ имеет наименьшую длину среди всех таких слов $\boldsymbol{\gamma}$, для которых $\boldsymbol{\alpha} \doteq \boldsymbol{\gamma\beta\delta}$, то вхождение $(\boldsymbol{\gamma_0}, \boldsymbol{\beta}, \boldsymbol{\delta_0})$ называется *первым вхождением слова* $\boldsymbol{\beta}$ *в* $\boldsymbol{\alpha}$. Двойственно определяется *последнее* вхождение $\boldsymbol{\beta}$ в $\boldsymbol{\alpha}$.

Слово $\boldsymbol{\beta}$ называется *началом* слова $\boldsymbol{\alpha}$ $(\boldsymbol{\alpha}, \boldsymbol{\beta} \in \boldsymbol{A}^\star)$, если $\boldsymbol{\alpha} \doteq \boldsymbol{\beta\delta}$ для подходящего слова $\boldsymbol{\delta}$ $(\boldsymbol{\delta} \in \boldsymbol{A}^\star)$. Слово $\boldsymbol{\beta}$ называется *концом* слова $\boldsymbol{\alpha}$, если $\boldsymbol{\alpha} \doteq \boldsymbol{\gamma\beta}$ для подходящего слова $\boldsymbol{\gamma}$ $(\boldsymbol{\gamma} \in \boldsymbol{A}^\star)$.

Замены и подстановки слов. Если $\boldsymbol{\alpha} \doteq \boldsymbol{\gamma\beta\delta}$, а $\boldsymbol{\alpha'} \doteq \boldsymbol{\gamma\beta'\delta}$ для некоторого слова $\boldsymbol{\beta'}$, то говорят, что *слово* $\boldsymbol{\alpha'}$ *получается из слова* $\boldsymbol{\alpha}$ *заменой вхождения* $(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\delta})$ *подслова* $\boldsymbol{\beta}$ *на слово* $\boldsymbol{\beta'}$. Пусть $\boldsymbol{\alpha}, \boldsymbol{\beta}$ — слова в некотором алфавите, а x — произвольная буква из этого алфавита. Заменим каждое вхождение буквы x в слово $\boldsymbol{\alpha}$ на слово $\boldsymbol{\beta}$. Получим новое слово, которое называется *результатом подстановки в слово* $\boldsymbol{\alpha}$ *вместо буквы x слова* $\boldsymbol{\beta}$. Его будем обозначать $(\boldsymbol{\alpha})_{\boldsymbol{\beta}}^{x}$.

УПРАЖНЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

1. Найдите все слова в алфавите $\{a,\, b\}$, имеющие:
(а) три (различных) непустых подслова;
(б) четыре непустых подслова.

2. Сколько непустых начал имеет слово $x_1x_2x_1x_1x_2$ в алфавите $\{x_1,\, x_2\}$?

3. Найдите все подслова слова $x_1x_2x_1x_2x_1x_2x_1$ в алфавите $\{x_1,\, x_2\}$, которые имеют в него точно два вхождения.

4. Второе вхождение слова xyx в слово $xyxyxyx$ замените на слово xx.

5. Найдите результат применения к слову $\boldsymbol{\alpha}$ следующих подстановок:
(а) $(abba)_{ba}^{b}$; (б) $(abba)_{ba}^{c}$;
(в) $(\boldsymbol{\alpha})_{\boldsymbol{\beta}}^{a}$, если $\boldsymbol{\alpha} \doteq aababb$, $\boldsymbol{\beta} \doteq bab$; (г) $(((xyxxyx)_z^x)_x^y)_y^z$.

6. С помощью одной подстановки вида $(\boldsymbol{\alpha})_{\boldsymbol{\beta}}^{v}$ слова вместо буквы получите из слова xy, если возможно, слово
(а) $zxyzxy$; (б) $zyxzyx$.

7. Слово $abbab$ получено из двух различных слов $\boldsymbol{\beta}$ и $\boldsymbol{\gamma}$ подстановкой слова ab вместо буквы c. Найдите эти слова. Сколько решений имеет данная задача?

8. Сколько всего существует таких слов $\boldsymbol{\alpha}$ в алфавите $\{x, y, z\}$, что $(\boldsymbol{\alpha})_y^z \doteq xyxyy$?

9. Из какого неоднобуквенного слова $\boldsymbol{\alpha} \in \{a,\, b\}^\star$ однократной подстановкой слова вместо некоторой одной буквы можно получит как слово $abbab$, так и слово $babba$?

◊ ◊ ◊

10. Найдите все слова наибольшей длины в алфавите $\{a,\, b\}$, такие, что в каждое из этих слов всякое подслово, длина которого больше единицы, входит не более одного раза.

11. Какое наибольшее количество различных подслов длины 2 может иметь слово, содержащее вхождения точно двух различных букв алфавита? Приведите пример такого слова наименьшей длины.

12. Докажите, что для всякого $n \geqslant 6$. существует слово $\boldsymbol{\alpha}$ в алфавите $\{a,\, b\}$ длины n, имеющее вхождения различных букв и такое, что после вычеркивания из него любого вхождения любой буквы, получается слово, имеющее то же множество подслов длины 2, что и слово $\boldsymbol{\alpha}$.

13. Найдите подслова слова $x_1x_2x_1x_2x_1x_2x_1$ в алфавите $\{x_1,\, x_2\}$, имеющие в него точно два вхождения и замените первое из этих вхождений на слово $x_1x_2x_1$. Сколько решений имеет эта задача?

14. Приведите пример такого слова $\boldsymbol{\alpha}$, что $(\boldsymbol{\alpha})_{ab}^{a} \doteq (\boldsymbol{\alpha})_{ba}^{b}$, в алфавитах $\{a,\, b\}$, $\{a,\, b,\, c\}$. Какова наименьшая длина такого слова?

15. Получите слово $babbbabb$ из слова abc, пользуясь, возможно несколькими, подстановками вида $(\boldsymbol{\alpha})_{\boldsymbol{\beta}}^{x}$, где $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \{a,\, b,\, c\}^\star$, $x \in \{a,\, b,\, c\}$.

16. Как несколькими подстановками вида $(\boldsymbol{\alpha})_{\boldsymbol{\beta}}^{x}$ из слова ab в алфавите $\{a,\, b,\, c\}$ получить слово ba? Возможно ли это сделать в алфавите $\{a,\, b\}$?

17. Покажите, что из слова $abab$ подстановками букв вместо букв невозможно получить ни одно из следующих слов:
(а) bab; (б) $babab$; (в) $abba$.

18. Можно ли из слова $x_1x_2x_3x_3x_0x_3$ получить подстановками букв вместо букв следующие слова:
(а) $x_2x_1x_2x_2x_1x_1$; (б) $x_2x_1x_2x_2x_1x_2$;
(в) $x_0x_0x_1x_2x_0x_2$; (г) $x_4x_5x_1x_1x_2x_1$.