

5. (8)

Даны условные частоты (частотности) встречаемости 18 фонем русского языка:

Фонема	Частотность
A	65
o	42
e	57
u	26
i	46
p	28
b	38
d	35
t	15
g	24
k	61
v	12
f	17
z	55
s	19
h	8
m	16
r	43

Требуется объединить фонемы в 3 группы (кластера) методом восходящей иерархической кластеризации на основе информации о частотности встречаемости фонем.

Введем понятие расстояния между кластерами:

Функция расстояния между кластерами $\rho(C_1, C_2)$ определена как сумма частотностей этих кластеров.

Алгоритм кластеризации:

1. Каждая фонема образует отдельный кластер.
2. Два самых близких кластера объединяются на основании расстояния между кластерами ρ (значение функции расстояния между кластерами наименьшее) и считаются одним новым кластером. Количество кластеров уменьшается на один. При объединении кластеров их частотности складываются.
3. Переход к шагу 2, пока не останется заданное число кластеров.

Перечислите в ответе число элементов каждого кластера после кластеризации по убыванию. Например, если количество элементов равно 1, 9, 8, то в ответе следует записать 981.

Отправить

Предыдущая

Следующая