

n-gramy w analizie sekwencji biologicznych

Michał Burdukiewicz¹, Piotr Sobczyk²

¹Zakład Genomiki, Uniwersytet Wrocławski

²Instytut Matematyki i Informatyki, Politechnika Wrocławska

Outline

1 n-gramy (k-mery)

n-gramy (k-mery, k-tuple) to wektory o długości n zawierające znaki z sekwencji wejściowych.

	P1	P2	P3	P4	P5	P6
S1	C	T	T	A	G	C
S2	C	A	G	A	C	G
S3	G	T	G	A	T	T

Przykładowe sekwencje. S - sekwencje, P - pozycja nukleotydu.

	A	C	G	T
S1	1	2	1	2
S2	2	2	2	0
S3	1	0	2	3

Zliczenia 1-gramów.

	P1	P2	P3	P4	P5	P6
S1	C	T	T	A	G	C
S2	C	A	G	A	C	G
S3	G	T	G	A	T	T

Przykładowe sekwencje. S - sekwencje, P - pozycja nukleotydu.

	AA	CA	GA	TA	AC	CC	GC	TC
S1	0	0	0	1	0	0	1	0
S2	0	1	1	0	1	0	0	0
S3	0	0	1	0	0	0	0	0

Zliczenia 2-gramów (fragment tabeli).