

Вспомогательное слово 'hotel'. Требуется сравнить $P(\text{hotel})$, $P(\text{hostel})$.

Пусть $P(\text{hostel}) = m$, тогда $P(\text{hotel}) = 5m$, т.к. слово 'hotel' встречается в 5 раз чаще. Пусть $\text{count}[s] = n$, знаем, $\text{count}[o] = 1,2n$.

1) Проверим seq. сравнение:

x	w	$x w$	$P(x w)$	$P(w)$
hotel	hotel	od o	$\text{ins}[o, od] / \text{count}[o]$	$5m$
hotel	hostel	d s	$\text{sub}[d, s] / \text{count}[s]$	m

$$\begin{aligned} \bullet P(\text{hotel}) &= P(\text{hotel} | \text{hotel}) \cdot P(\text{hotel}) = 5m \cdot \frac{\text{ins}[o, od]}{\text{count}[o]} = \\ &= 5m \cdot \frac{9}{1,2n} = 37,5 \frac{m}{n} \end{aligned}$$

$$\begin{aligned} \bullet P(\text{hostel}) &= P(\text{hotel} | \text{hostel}) \cdot P(\text{hostel}) = m \cdot \frac{\text{sub}[d, s]}{\text{count}[s]} = \\ &= m \cdot \frac{7}{n} = 7 \frac{m}{n} \end{aligned}$$

$$\bullet \frac{P(\text{hotel})}{P(\text{hostel})} = \frac{37,5 \frac{m}{n}}{7 \frac{m}{n}} = 5,35 > 1 \Rightarrow \text{В паре seq. сравнения слово 'hotel' более вероятно.}$$

2) Проверим w. сравнение по длине:

Пусть A - алфавит символов

x	w	x w	P(x w)	P(w)
hwdtel	htel	od o	$(ins[0, od] + 1) / (count[o] + A)$	5m
hwdtel	hostel	d s	$(sub[d, s] + 1) / (count[s] + A)$	m

$$\bullet P(htel) = P(hwdtel | htel) \cdot P(htel) = 5m \cdot \frac{ins[0, od] + 1}{count[o] + A} =$$

$$= 5m \cdot \frac{9+1}{1,2n+A} = \frac{50m}{1,2n+A}$$

$$\bullet P(hostel) = P(hwdtel | hostel) \cdot P(hostel) = m \cdot \frac{sub[d, s] + 1}{count[s] + A} =$$

$$= m \cdot \frac{7+1}{n+A} = \frac{8m}{n+A}$$

$$\bullet \frac{P(htel)}{P(hostel)} = \frac{\frac{50m}{1,2n+A}}{\frac{8m}{n+A}} > 1 \Rightarrow \text{В паріме в снмубарнем даннаца нубо 'htel' донее бепараму.}$$