Cílem dnešního cvičení je pochopit, proč je kukaččí hashování dobré. Budeme předpokládat, že máme k dispozici „dokonalou rodinu hashovacích funkcí $\mathcal{H}$“, pro kterou platí

$$\Pr_{h \in \mathcal{H}}[h(x) = i] = \frac{1}{m},$$

kde m je velikost oboru hodnot právě voleného h. Zpravidla je $\operatorname{rng}(h) = \{0, \dots, m-1\}$.

Definice. Mějme kukaččí hashovací tabulku velikosti m obsahující prvky S, kde $n = |S|$. *Kukaččí graf* G má vrcholy $V(G) = \{0, \dots, m-1\}$. Pro každý prvek $x \in S$ obsahuje G neorientovanou hranu $\{h_1(x), h_2(x)\}$.

Užitečné nástroje.

- **Union bound.** Nechť $A_1, \dots, A_n$ jsou nějaké jevy. Pak $\Pr[A_1 \cup \dots \cup A_n] \leq \Pr[A_1] + \dots + \Pr[A_n]$. (Uměli byste důkaz?)
- **Linearita střední hodnoty.** Nechť $X_1, \dots, X_n$ jsou nějaké náhodné veličiny a $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. Pak $\mathbb{E}[\alpha_1 X_1 + \dots + \alpha_n X_n] = \alpha_1 \mathbb{E}[X_1] + \dots + \alpha_n \mathbb{E}[X_n]$.
- **Podmíněná pravděpodobnost.** Pravděpodobnost $\Pr[A \mid B]$ (čteno „pravděpodobnost jevu A za předpokladu, že nastal jev B“) je $\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}$, neboli $\Pr[A \cap B] = \Pr[A \mid B] \Pr[B]$.

1. Kdy může dojít k selhání vkládání? To se stane, pokud pro vkládaný prvek x platí, že $h_1(x)$ leží na cyklu kukačkového grafu.

   Dokažte následující tvrzení.

   Nechť $c > 1$ je konstanta a $\frac{n}{m} \leq \frac{1}{2c}$. Pro dva vrcholy $s, t \in V(G)$ platí, že pravděpodobnost existence cesty z s do t délky k je nejvýše $\frac{1}{mc^k}$.

   Pro řešení následujících úloh můžete předpokládat, že jste vyřešili Úlohu 1.

2. Jak dlouho může trvat vkládání prvku x?

   Dokažte následující tvrzení.

   Nechť $c > 1$ je konstanta. Střední délka cesty začínající v $h_1(x)$ je $\mathcal{O}(1)$.

3. Kolikrát může dojít k přehashování?

   Dokažte následující tvrzení.

   Nechť $c > 1$ je konstanta. Střední počet přehashování při vkládání n prvků do prázdné tabulky velikosti m je $\mathcal{O}(1)$, pokud $\frac{n}{m} \leq \frac{1}{2c}$.

Poznámka. Kukačkové hashování vyžaduje, aby tabulka byla zaplněná z maximálně 25 %. Jde získat konstantní operace (s vysokou pravděpodobností) a mít zaplněnost, dejme tomu, aspoň 99 %? Ano,

`https://arxiv.org/abs/2109.04548v2`

Poznámka. Mít zcela dokonalou rodinu hashovacích funkcí je nepraktické. Zkoumat vlastnosti rodin hashovacích funkcí je naopak otravné (např $\{ax \bmod m\}_{a \in \mathbb{N}}$ je 2-nezávislé). Stejně používáme náhodné bity, je možné je přímo použít na konstrukci slovníků? Ano,

`https://arxiv.org/abs/2209.06038`

Užitečná poznámka. Až budete implementovat kukaččí hashování v domácím úkolu, nevymýšlejte blbosti a pseudokód vkládání prostě opište. Sice se vám bude zdát, že dochází k nekonečné rekurzi, ale to je ok.