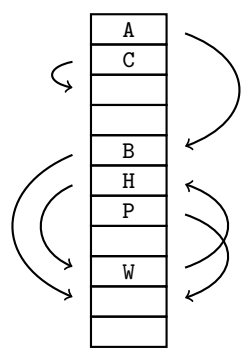
Análise

Nós já vimos que as operações de busca e remoção executam em tempo $O(1)$, no pior caso.

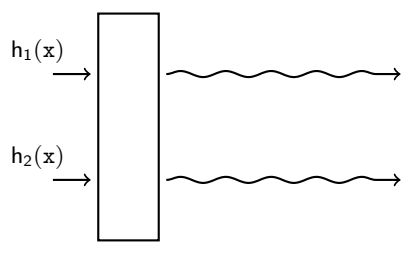
Mas, qual é o tempo médio da operação de inserção ?

Bom, para responder essa pergunta, é útil considerar o *grafo de vizinhanças* da tabela cuco



Quer dizer, nesse esquema as setas indicam para onde o elemento vai, quando ele é tirado da sua posição.

Daí, não é difícil ver que a busca percorre dois caminhos nesse grafo



E o tempo médio da inserção é proporcional ao tamanho dos caminhos que ela percorre.

Em outras palavras, a operação de inserção é eficiente se os caminhos que surgem no grafo de vizinhanças da tabela cuco são curtos.

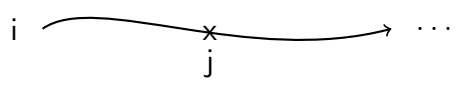
A seguir, nós vamos ver que se a tabela contém muitas posições vazias, então os caminhos são curtos e a inserção executa em tempo médio $O(1)$.

Especificamente, considere uma tabela de tamanho n.

E suponha que ao menos $2n/3$ posições estão vazias.

Para determinar o tamanho médio dos caminhos que surgem nessa tabela, nós fazemos a seguinte pergunta

- *Qual é a probabilidade de que o caminho que começa na posição i passe pela posição j ?*



Bom, a situação mais simples possível é aquela onde existe uma seta de i para j.

Isso acontece quando existe algum elemento x na tabela tal que

$$\mathsf{h}_1(\mathsf{x}) \;=\; \mathsf{i} \qquad \text{e} \qquad \mathsf{h}_2(\mathsf{x}) \;=\; \mathsf{j}$$

Mas a probabilidade de que isso aconteça para um certo elemento x é igual a

$$\frac{1}{\binom{n}{2}} \;=\; \frac{2}{n(n-1)}$$

— (*assumindo que as funções* $\mathsf{h}_1, \mathsf{h}_2$ *tem comportamento aleatório*)

Por outro lado, existem até $n/3$ elementos na tabela.

Mas ainda assim, a probabilidade de que algum desses elementos coloque uma seta de i para j é no máximo

$$\frac{n}{3} \cdot \frac{2}{n(n-1)} \;\simeq\; \frac{2}{3n}$$

— (*o que é bem pequenininho, quando n é bem grande ...*)

A seguir, nós podemos considerar a possibilidade de que exista um caminho de tamanho 2 entre i e j

$$\mathsf{i} \longrightarrow \mathsf{k} \longrightarrow \mathsf{j}$$

Como indicado no esquema, para que isso aconteça é preciso que exista uma posição k tal que

- existe uma seta entre i e k
- existe uma seta entre k e j

Mas, nós acabamos de ver que a probabilidade de uma seta qualquer é de $2/3n$.

Daí que, a probabilidade de que as duas setas acima existam, para uma certa posição k, é igual a

$$\frac{2}{3n} \cdot \frac{2}{3n} \;=\; \left(\frac{2}{3}\right)^2 \cdot \frac{1}{n^2}$$

Por outro lado, existem n posições na tabela que podem fazer o papel de k.

Mas ainda assim, a probabilidade de que exista uma caminho de tamanho 2 entre i e j é no máximo

$$n \cdot \left(\frac{2}{3n}\right)^2 \cdot \frac{1}{n^2} \;=\; \left(\frac{2}{3}\right)^2 \cdot \frac{1}{n}$$

— (*o que é bem pequenininho, quando n é bem grande ...*)

Agora já deu para o ver o padrão, não é?

Quer dizer, a probabilidade de que exista um caminho de tamanho 3 entre i e j é

$$\left(\frac{2}{3}\right)^3 \cdot \frac{1}{n}$$

e a probabilidade de um caminho de tamanho 4 é de

$$\left(\frac{2}{3}\right)^4 \cdot \frac{1}{n}$$

Daí, a gente observa que

$$\begin{aligned} P\big(\text{existe um caminho entre i e j}\big) \;=\; & P\big(\text{existe caminho de tamanho 1}\big) \\ + \; & P\big(\text{existe caminho de tamanho 2}\big) \\ + \; & P\big(\text{existe caminho de tamanho 3}\big) \\ + \; & P\big(\text{existe caminho de tamanho 4}\big) \\ + \; & \ldots \end{aligned}$$

E agora nós podemos escrever

$$P\big(\text{existe um caminho entre i e j}\big) \;=\; \left(\frac{2}{3}\right) \cdot \frac{1}{n} \;+\; \left(\frac{2}{3}^2\right) \cdot \frac{1}{n} \;+\; \left(\frac{2}{3}^3\right) \cdot \frac{1}{n} \;+\; \ldots$$

o que a gente reescreve assim

$$P\big(\text{existe um caminho entre i e j}\big) \;\leq\; \frac{1}{n} \cdot \sum_{k=1}^{\infty} \left(\frac{2}{3}\right)^k$$

e depois assim

$$P\big(\text{existe um caminho entre i e j}\big) \;\leq\; \frac{1}{n} \cdot \frac{2/3}{1 - 2/3} \;=\; \frac{2}{n}$$

— (*o que é bem pequenininho, quando n é bem grande ...*)

Quer dizer, a probabilidade de que a posição j apareça no caminho que começa em i é bem pequena.

E isso vale para qualquer posição tabela.