



11-711 Recitation

Nikita Srivatsan

2017-09-15

BuzzFeed

[News](#)

[Videos](#)

[Quizzes](#)

[Tasty](#)

[More](#) ▾

Search



CELEBRITY

7 Neat Tricks for Improving Memory Usage and Computation Speed

NUMBER 3 WILL CHANGE YOUR ASYMPTOTIC COMPLEXITY!!!

Posted on September 14, 2017, at 3:53 p.m.

BuzzFeedNEWS

Donald Trump insisted nothing is final on a DREAMer deal with Democrats, but said, "We're working on a plan for DACA." >

A student was killed in a school shooting near Spokane, Washington, as he tried to stop the gunman. Police detained a suspect. >

Data Types

Max trigram count **468,261***

Max bigram count **7,109,704***

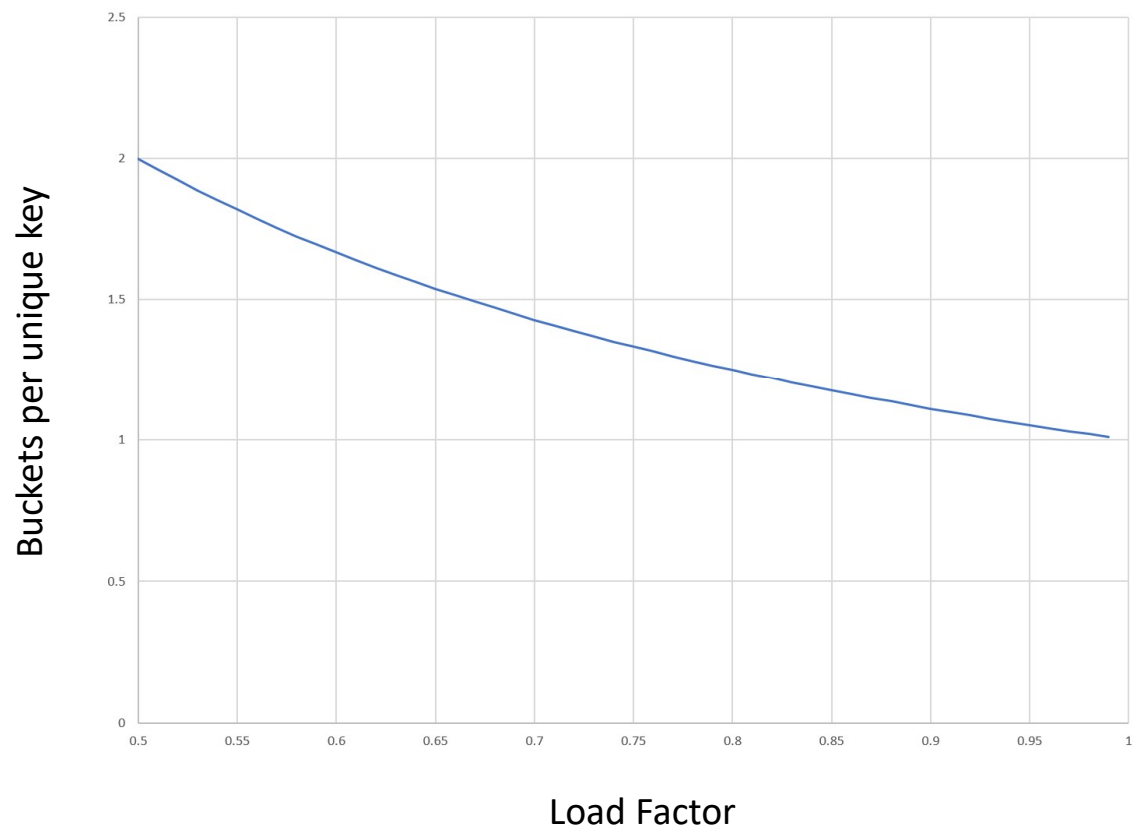
Max unigram count **19,880,264**

N_{1+} counts? **$\leq 495,172$**

MAX_INT **2,147,483,647**

MAX_SHORT **32,767**

Load Factor - Memory



Load Factor - Speed

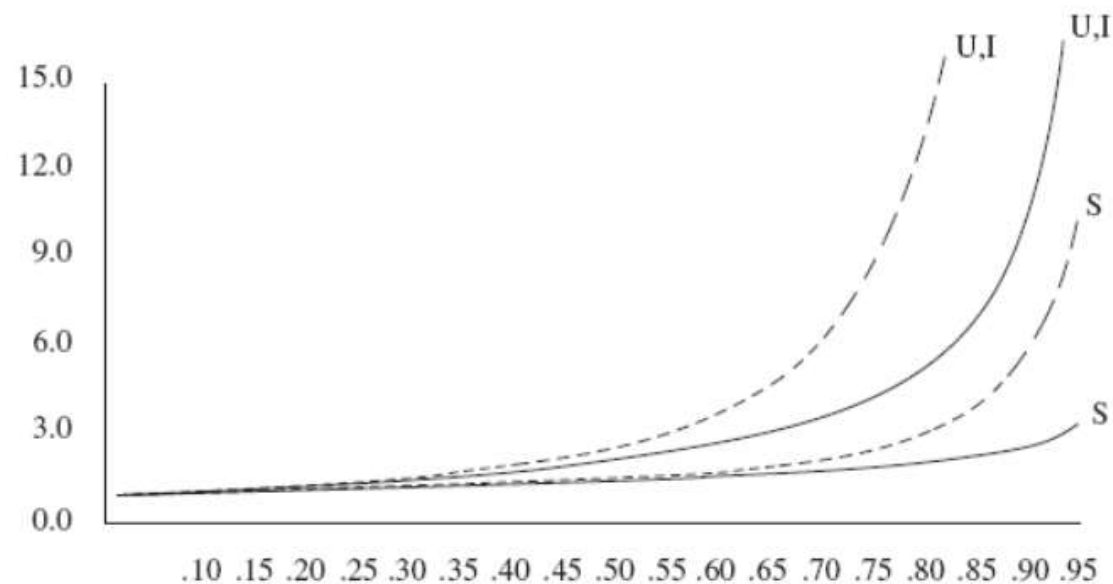


Figure 5.12 Number of probes plotted against load factor for linear probing (dashed) and random strategy (*S* is successful search, *U* is unsuccessful search, and *I* is insertion)

Image Credit: Chris Brown

Hash Functions

```
key = key ^ (key >> 32)
```

01101011

00000110

0110**1101**

Avoiding Loops

$$\begin{aligned}
 P(w_3|w_1w_2) &= \frac{\max(c(w_1w_2w_3) - d, 0)}{\sum_{v \in V} c(w_1w_2v)} + \alpha(w_1w_2)P(w_3|w_2) = \\
 &= \frac{\max(c(w_1w_2w_3) - d, 0)}{c(w_1w_2)} + \alpha(w_1w_2)P(w_3|w_2)
 \end{aligned} \tag{5}$$

$$\alpha(w_1w_2) = d \cdot \frac{N_{1+}(w_1w_2\bullet)}{c(w_1w_2)} \tag{10}$$

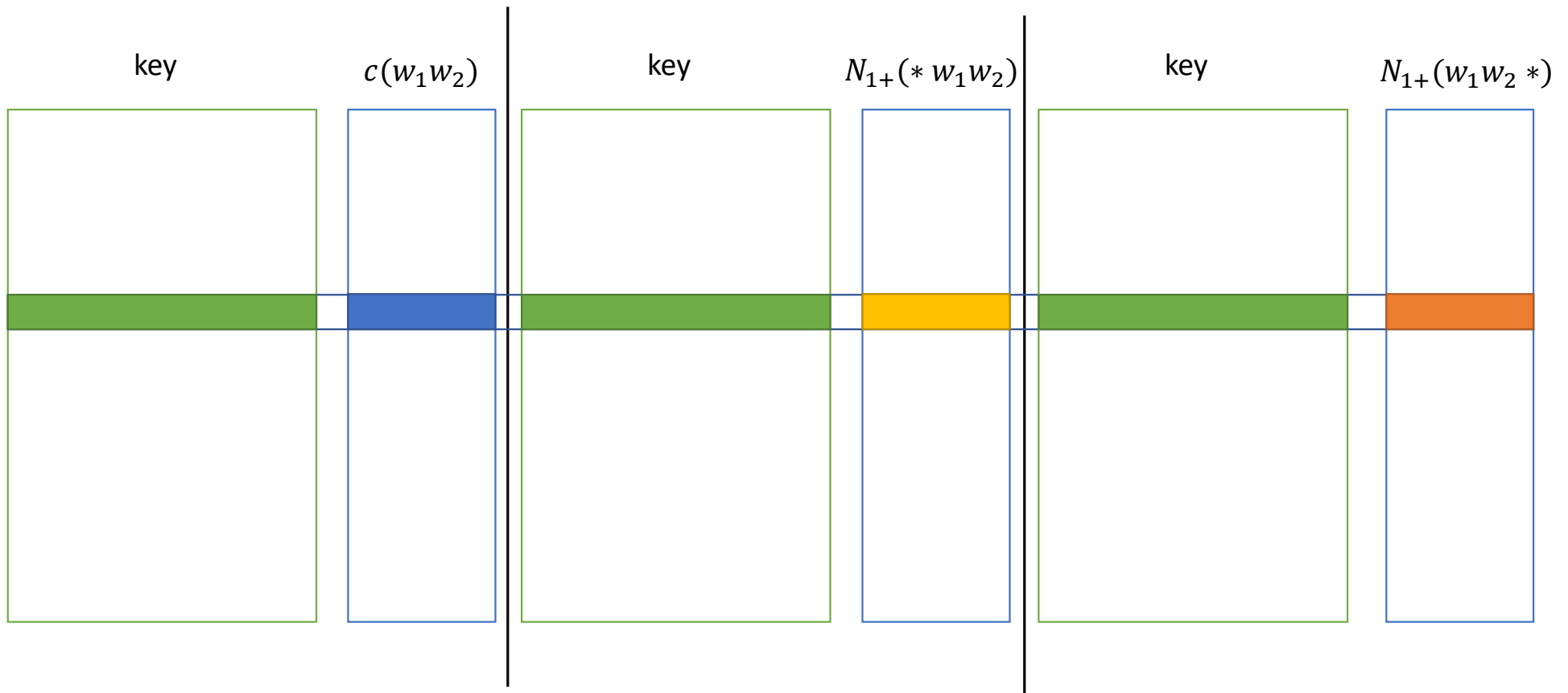
$$\begin{aligned}
 P(w_3|w_2) &= \frac{\max(N_{1+}(\bullet w_2w_3) - d, 0)}{\sum_{v \in V} N_{1+}(\bullet w_2v)} + \alpha(w_2)P(w_3) = \\
 &= \frac{\max(N_{1+}(\bullet w_2w_3) - d, 0)}{N_{1+}(\bullet w_2\bullet)} + \alpha(w_2)P(w_3)
 \end{aligned} \tag{6}$$

$$\alpha(w_2) = d \cdot \frac{N_{1+}(w_2\bullet)}{N_{1+}(\bullet w_2\bullet)} \tag{13}$$

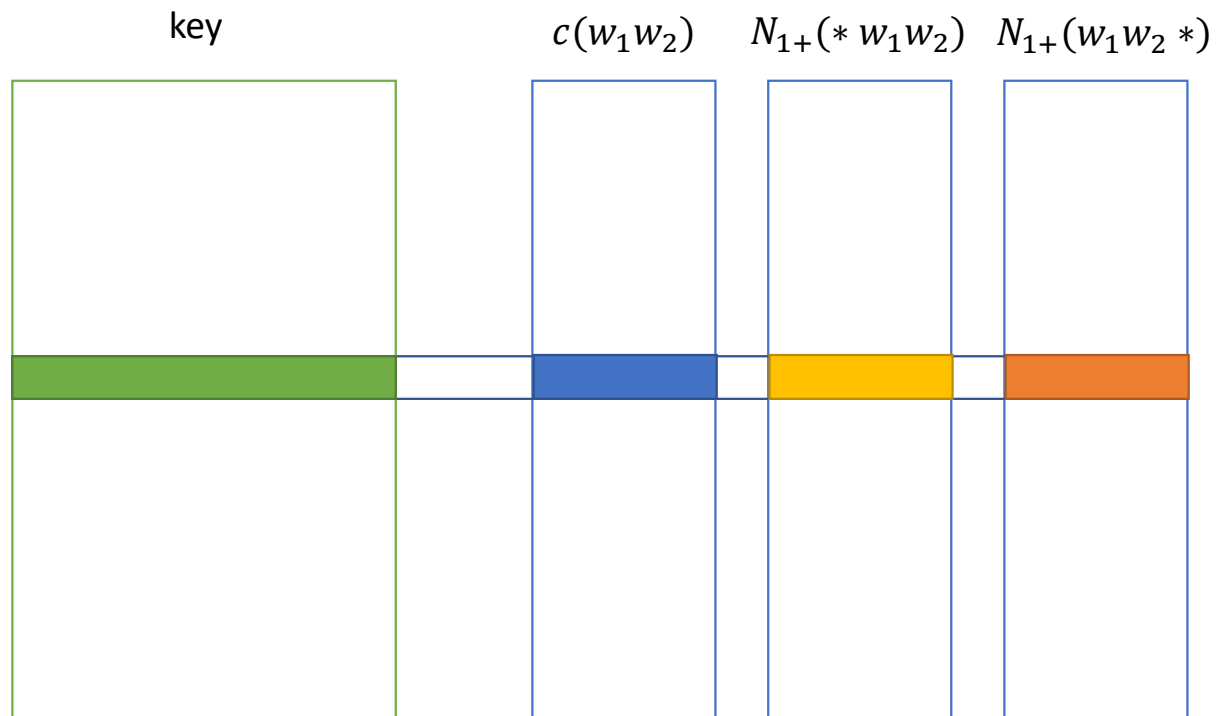
$$P(w_3) = \frac{N_{1+}(\bullet w_3)}{\sum_{v \in V} N_{1+}(\bullet v)} = \frac{N_{1+}(\bullet w_3)}{N_{1+}(\bullet\bullet)} \tag{7}$$

Image Credit: Maria Ryskina

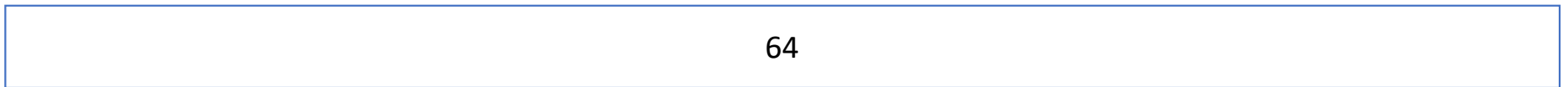
Multi-Value Hash Tables



Multi-Value Hash Tables



Byte Arrays (tentative)



Caching

$w_1 w_2 w_3$

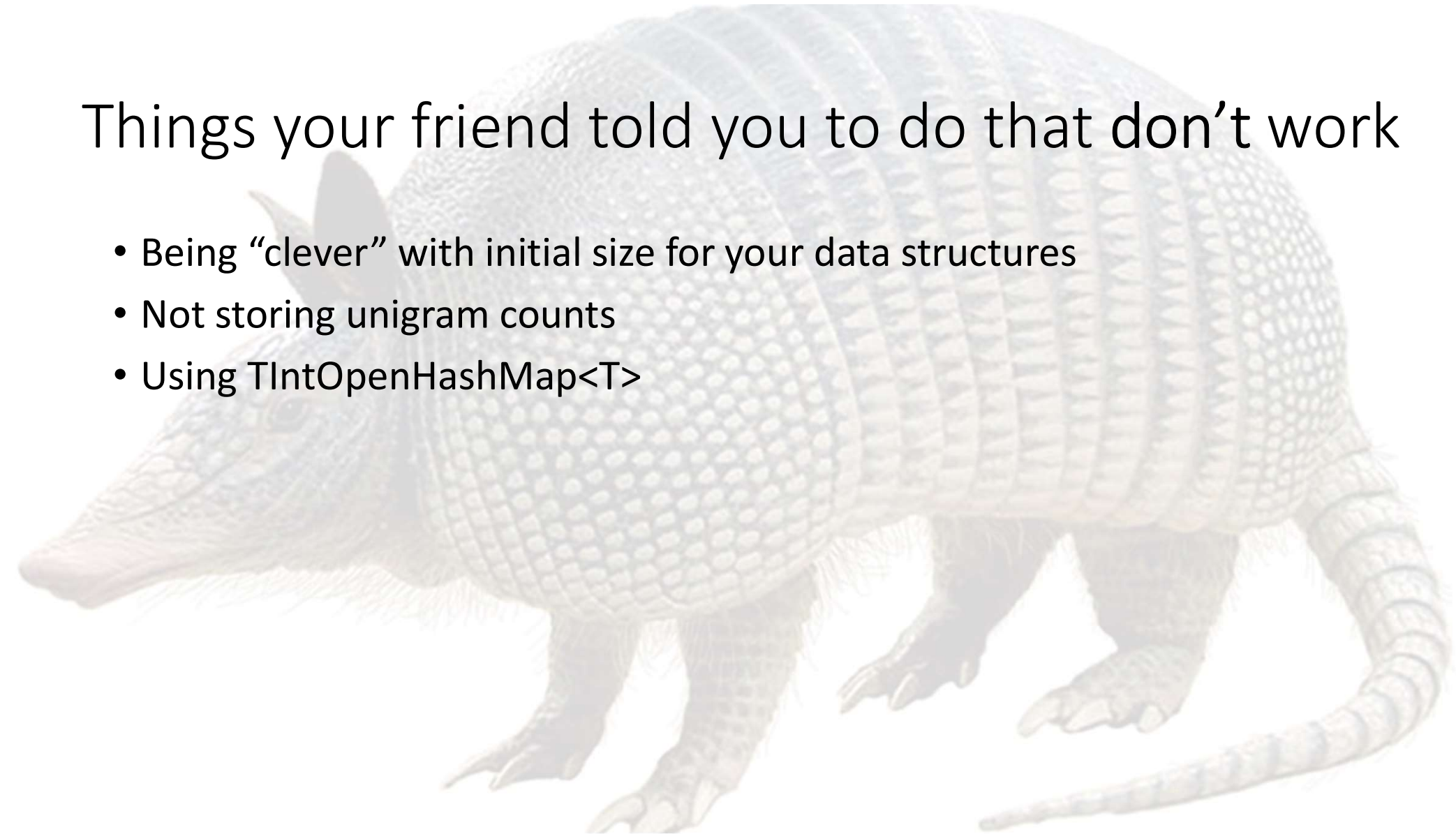


$P(w_3 | w_1 w_2)$



Things your friend told you to do that don't work

- Being “clever” with initial size for your data structures
- Not storing unigram counts
- Using `TIntOpenHashMap<T>`



Questions?