March 16 – Hashing
Wade Fagen-Ulmschneider
A Hash Table based Dictionary

Client Code:

```
1 | Dictionary<KeyType, ValueType> d;
2 | d[k] = v;
```

A **Hash Table** consists of three things:
1. A hash function, $f(k)$
2. An array
3. A mystery element
A Perfect Hash Function

(Angrave, CS 241)
(Beckman, CS 421)
(Cunningham, CS 210)
(Davis, CS 101)
(Evans, CS 126)
(Fagen-Ulmschneider, CS 225)
(Gunter, CS 422)
(Herman, CS 233)
A Perfect Hash Function

Keyspace: Rolling 5 dice!
Hash Function

Our hash function consists of two parts:

• A hash:

• A compression:

Choosing a good hash function is tricky…

• Don’t create your own (yet*)

• Very smart people have created very bad hash functions
Hash Function

Characteristics of a good hash function:

1. Computation Time:

2. Deterministic:

3. Satisfy the SUHA:
General Purpose Hash Function

Easy to create if: $|KeySpace| \sim N$
General Purpose Hash Function

Easy to create if: $|\text{KeySpace}| \sim N$

Difficult to Create:
General Purpose Hash Function

Easy to create if: $|\text{KeySpace}| \sim N$

Difficult to Create:
Given: Easy to create a hash function of strings of length 8.
Idea: Map 40 character things to length 8:

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book,' thought Alice 'without pictures or conversations?' So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her. There was nothing so very remarkable in that; nor did Alice think it so very much out of the way to hear the Rabbit say to itself, 'Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to ha
Idea: Map 40 character things to length 8:

https://en.wikipedia.org/wiki/Main_Page
https://en.wikipedia.org/wiki/2017_Austr
https://en.wikipedia.org/wiki/19th_Natio
https://en.wikipedia.org/wiki/Japanese_g
Hash Function

In CS 225, we focus on general purpose hash functions.

Other hash functions exist with different properties (e.g., cryptographic hash functions)
Collision Handling: Separate Chaining

$S = \{ 16, 8, 4, 13, 29, 11, 22 \}$ \quad |S| = n
$h(k) = k \% 7$ \quad |Array| = N

(Example of open hashing)
Collision Handling: Probe-based Hashing

\[ S = \{ 16, 8, 4, 13, 29, 11, 22 \} \quad |S| = n \]
\[ h(k) = k \mod 7 \quad |Array| = N \]

Try \( h(k) = (k + 0) \mod 7 \), if full…
Try \( h(k) = (k + 1) \mod 7 \), if full…
Try \( h(k) = (k + 2) \mod 7 \), if full…
Try …