#26: Hashing
October 25, 2017

Goals for Understanding Hashing:
1. We will define a **keyspace**, a (mathematical) description of the keys for a set of data.

2. We will define a function used to map the **keyspace** into a small set of integers.

A hash table consists of three things:
1. 
2. 
3. 

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)

...characteristics of this function?

A Second Hash Function

...characteristics of this function?

<table>
<thead>
<tr>
<th>Dictionary ADT in Client Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dictionary&lt;KeyType, ValueType&gt; d;</td>
</tr>
<tr>
<td>d[k] = v;</td>
</tr>
</tbody>
</table>

All hash functions will consist of two parts:
- A **hash**:
- A **compression**:
Characteristics of a good hash function:

1. Computation Time:
2. Deterministic:
3. SUHA:

Towards a general-purpose hashing function:
It is easy to create a general-purpose hashing function when the keyspace is proportional to the table size:
- Ex: Professors at CS@Illinois
- Ex: Anything you can reason about every possible value

It is difficult to create a general-purpose hashing function when the keyspace is large:

<table>
<thead>
<tr>
<th>Alice in Wonderland, Page 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 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 have looked further for it, and that she must have been very much mistaken, for, as she thought, 'I shall have to speak much sooner than I expected, if I am late.')</td>
</tr>
</tbody>
</table>

Reflections on Hashing
In CS 225, we are starting the study of general-purpose hash functions. There are many other types of hashes for specific uses:
- **Common**: Cryptographic hash functions

Even if we build a good hash function, it is not perfect. What happens when the function isn’t always a bijection?

CS 225 – Things To Be Doing:

1. Exam #8 (programming, MP4-like and AVL) starts Monday
2. MP5 is available now; extra credit +7 deadline is Monday, Oct. 30
3. lab_btree due on Sunday (Oct. 29)
4. Daily POTDs

Ex: Hashing a string of 8 characters is considered easy, 40 is not.