Teach A level Computing: Algorithms and Data Structures

Eliot Williams

@MrEliotWilliams

Queen Mary
University of London

SUPPORTED BY
MAYOR OF LONDON

COMPUTING AT SCHOOL
EDUCATE · ENGAGE · ENCOURAGE

KING'S
College
LONDON

Course Outline

1	Representations of data structures: Arrays, tuples, Stacks, Queues,Lists
2	Recursive Algorithms
3	Searching and Sorting - EW will be late!
4	Hashing and Dictionaries, Graphs and Trees
5	Depth and breadth first searching ; tree traversals

Queen Mary
University of London

SUPPORTED BY
MAYOR OF LONDON

COMPUTING AT SCHOOL
EDUCATE · ENGAGE · ENCOURAGE

KING'S
College
LONDON

Finding things quickly

Hash functions

tells us where in the entry to look. And so instead of having to look through the whole index, the hash function will tell us where that entry belongs.
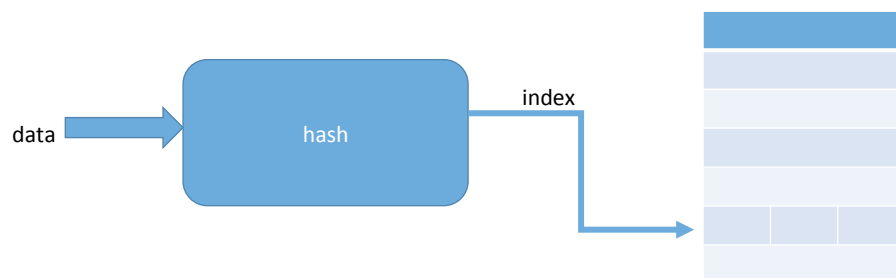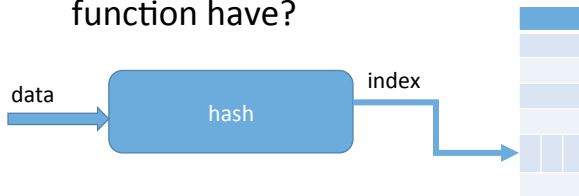
Thinking about hash functions

- we have b buckets in our hash table,
- and we have n items, and we should assume that n is much greater than b

(more items than buckets)

What properties should the hash function have?

A. [ ] Output a unique number between 0 and n-1

B. [ ] Output a number between 0 and b-1

C. [ ] Map approximately n/b keywords to bucket 0

D. [ ] Map approximately n/b keywords to bucker b-1

E. [ ] Map more keywords to bucket 0 than to bucket 1

data → hash → index

Defining a hash function

string →

b → Hash function → a number between 0 and b-1

Activity: Bad hash

string		
b	Bad Hash function	a number between 0 and b-1

- Define a procedure badHash
- that has two input parameters
  - A string representing the data item
  - The number of buckets
- It should return a bucket based on the first letter in the keyword.
  - modulo the number of buckets!

Bad hash

```
def badhash(word,buckets):
    return ord(word[0])%buckets
```

Why is bad hash bad?

- [ ] It takes too long to compute
- [ ] It produces an error for one input keyword
- [ ] If the keywords are distributed like words in English, some buckets will get too many words
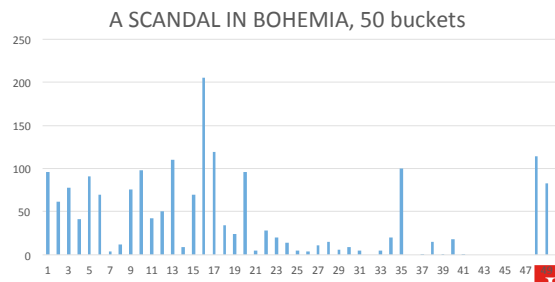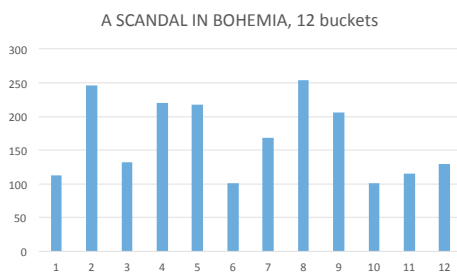- [ ] If the number of buckets is large, some buckets will not get any keywords

Testing hash quality

Bad hash

- For SIR ARTHUR CONAN DOYLE'S "A Scandal in Bohemia" the distribution of words to buckets from Badhash is

A SCANDAL IN BOHEMIA, 12 buckets

A SCANDAL IN BOHEMIA, 50 buckets

Better hash

- To ensure a more even distribution we don't want all the words beginning with the same letter to end up in the same bucket.
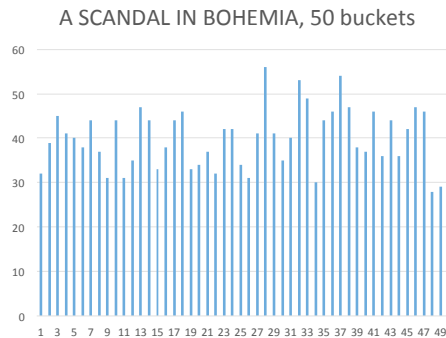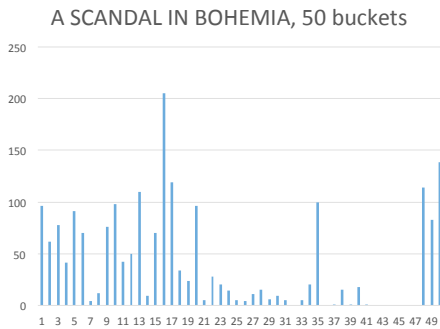
- How could we do this

-

Bad vs better hash

Perfect hash function

- A perfect hash function distributes keys evenly across all the buckets.

- Which of the following will leave the expected look up time for a given keyword essentially unchanged.
- [ ] Double number of keywords, same # of buckets
- [ ] Same number of keywords, double # of buckets
- [ ] Double number of keywords, double # of buckets
- [ ] Halve number of keywords, same # of buckets
- [ ] Halve number of keywords, halve # of buckets

Key words

- **Hash table**:
- The table gets its name from the method used to determine the row to use. The hash value generated by applying the hash function to the key is the table index where the record should be stored if the row is free.

- **Hash function**:
- Is a function H, applied to a key k, which generates a hash value $H(k)$ of range smaller than the domain of values of k.

- **Hash value**:
- is the value generated by the application of the hash function to the key. This value can be used as the hash table index where the key and any associated data can be stored if this table location is free.
- **Hash key**:
- is the key that the hash function is applied to
- **Hashing**:
- The process of applying a hash function to a key to generate a hash value

Entering and retrieving data into a hashtable

- Given a (*8-digit*) ULN number *k*, its **hash value** *H*(*k*) is calculated as follows using the **hash**
- **function** *H*, where
- **H(k) = k Mod b**

Thinking about collisions

- A collision occurs when two or more different keys hash to the same value
- Collisions can be resolved in two ways:
- 1 *open addressing/closed hashing* Store the record in the "next available" location in the table,
- 2 *open hashing* Store a pointer in each table location that points to a list of records that have all collided at this table location otherwise set the pointer value to null.
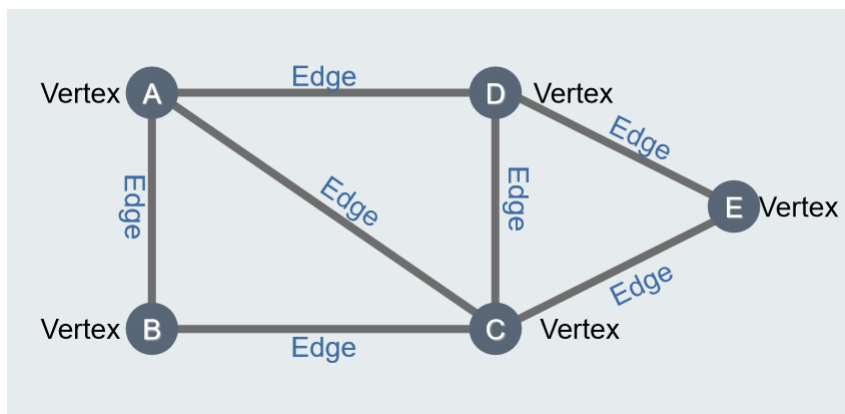
Dictionaries in python

Graphs

Queen Mary
University of London

SUPPORTED BY
MAYOR OF LONDON

COMPUTING AT SCHOOL
EDUCATE · ENGAGE · ENCOURAGE

KING'S
College
LONDON

What is a graph

Queen Mary
University of London

SUPPORTED BY
MAYOR OF LONDON

COMPUTING AT SCHOOL
EDUCATE · ENGAGE · ENCOURAGE

KING'S
College
LONDON

Terminology

- A graph can have cycles (loops)
- A weighted graph has a value on the edge
- A directed graph has arrows to indicate direction
- A multigraph has multiple edges connecting the same vertices
- A tree is a type of graph where…
  - All vertices are connected
  - The graph is undirected
  - There are no cycles