This week in Compsci 6/101

- **From sets to dictionaries**
  - ➤ **Incredibly powerful in solving problems**
  - ➤ **Power in Perl, Python, …**

- **Work on APTs, toward Hangman and Jotto**
  - ➤ **More on existing APTs, then mastery APTs**
  - ➤ **What do APTs teach you and us?**

- **More on communication between modules**
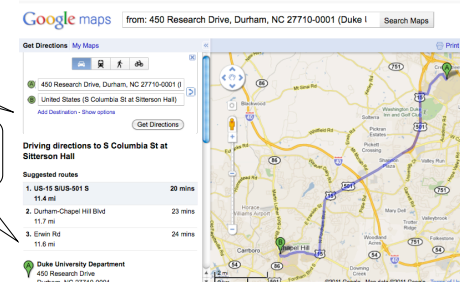 - ➤ **Organizing code to create programs**

Useful Computer Science

```
http://maps.google.com/maps?f=d&source=s_d&saddr=450+Research+Drive,
+Durham,+NC+27710-0001+(Duke+University+Department)&daddr=United+States
+(S+Columbia+St+at+Sitterson+Hall)&hl=en&geocode=FbJnJQIdGXZL-
yGNoxzuVRYfsg%3BFUbxIwIdkrxJ
yGmmUTWHCQpCQ&mra=pd&sll=35.957876,-78.99716&sspn=0.145342,0.200844&ie=
UTF8&ll=35.958833,-78.997192&spn=0.145341,0.200844&z=12
```
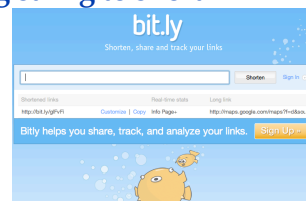
Sitterson Hall, Chapel Hill, NC

Research Drive, Durham, NC

How does this work?

- **http://bit.ly/glFvFi**
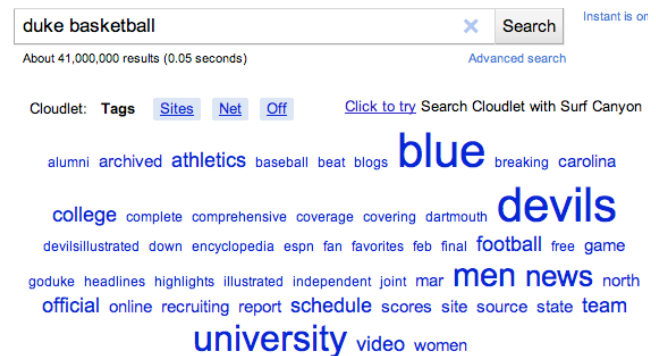 - ➤ **URL Shortener, why useful on Twitter?**
  - ➤ **How are URLs stored, searched for, used?**

- **Hashing: convert long string to short number**
  - ➤ **glFvFi is a number!**
  - ➤ **Look up and use!**

Count word occurrences: tag cloud

How do T9 and iTap work?

- **See wikipedia entries**
  - predict as you type
  - modify user db

- **What's stored on phone?**
  - why is a patent relevant?

- **How many words map to 752837? others?**
  - How do you write code to figure this out
  - How do you search for keypresses efficiently?

Maria Klawe

Chair of Computer Science at UBC, Dean of Engineering at Princeton, *President of Harvey Mudd College*, ACM Fellow,...

Klawe's personal interests include painting, long distance running, hiking, kayaking, juggling and playing electric guitar. She describes herself as "crazy about mathematics" and enjoys playing video games.

"I personally believe that the most important thing we have to do today is use technology to address societal problems, especially in developing regions"

What is a Literary Fingerprint?

- http://www.physorg.com/news179651371.html
- http://iopscience.iop.org/1367-2630/11/12/123015

- **What are some of the issues in creating 'fingerprint'**
  - Where else do fingerprints occur?
  - What is www.shazam.com
  - What is www.tineye.com

- **How do we go from sets to frequency analysis?**
  - Understanding Python *dictionary* data type

Literary Fingerprint

- **Timing and playing with code in fingerPrint.py**
  - How do we find out how fast this is?
  - How do we change the format of the output?
  - Can we organize output differently?

- **How can we find 'other' fingerprints**
  - Shazaam, genome, images
  - What will be the key, what will be the value?

How do you read words from file?

- **I exist solely as the *Canis Familiaris 'Bassett'***
- **I ain't nothin' but a hound-dog**

What does this code do? How?

```
def fileStatsList(filename):
    file = open(filename)
    stats= []
    for word in file.read().split():
        found = False
        for pair in stats:
            if pair[0] == word:
                pair[1] += 1
                found = True
                break
        if not found:
            stats.append([word,1])
    file.close()
    return stats
```

Faster, Cheaper, Totally in Control

```
def fileStatsDictionary(filename):
    file = open(filename)
    stats = {}
    for word in file.read().split():
        if word in stats:
            stats[word] += 1
        else:
            stats[word] = 1
    file.close()
    return stats
```

Accessing Python dictionaries

- **Index dictionary with key, using [...]**
  - ➤ **Like indexing a list, but *index* is a string**
  - ➤ **Internally creates number from key**
    - • **This association is known as *hashing* the key**
    - • **Key must be immutable so number doesn't change!**
- **Methods for dictionaries:**
  - ➤ **.clear(), .get(key), .get(key,default)**
  - ➤ **.keys(), .values(), .items()**

- **When using iteration or x in d, we're talking keys**
  - ➤ **Iterate over keys, query about keys, and so on**