

Plan for the Week

- **Review Dictionaries and Set operations**
 - In the context of solving APTs
 - See also `.join` and either `.sort()` or `sorted(seq)`
 - Note that `.join()` and `sorted(seq)` return values
 - Note that `.sort()` does NOT return a value, therefore ...
- **Collaborative Filtering/Recommender Assignment**
 - How does Amazon know what I like? Recommended?
 - What does Netflix think I want to watch?
- **Sorting in Python (essential for recommender)**

When to stop iterating?

- **Definite loops are "simple"**
 - `for elt in collection: ...`
 - `for i in range(a,b): ...`
- **What about indefinite loops?**
 - `while attempt < max_allowed: ...`
 - `while True: ... break ...`
- **Example: Thesaurus APT, InternetSecurity APT**
 - When to stop iterating? When a condition is true
 - `ok = False, ... compute ... ok = True ... if ok: break`

Python Control Review

- **How to break out of loop? How to continue?**
 - `break` and `continue`
- **Loop over indexes, loop over values**
 - How to recognize which approach to use?
 - Combo: `for index, value in enumerate(seq):`

```
z = ['dog', 'cat', 'bear']
for ind,val in enumerate(z):
    print "%s\t%s" % (ind,val)
```

```
0 'dog'
1 'cat'
2 'bear'
```

Eugene (Gene) Myers

- Lead computer scientist/
software engineer at Celera
Genomics, then at Berkely,
now at Janelia Farms Research
Institute (HHMI)



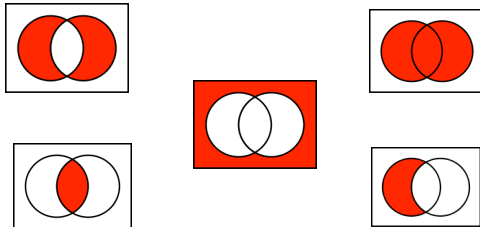
"What really astounds me is the architecture of life. The system is extremely complex. It's like it was designed." ... "There's a huge intelligence there."

- **BLAST and WG-Shotgun**

Patterns in some APTs

- **PositiveID and MemberCheck APTs**

- Set operations: $|$, $\&$, $-$, $\wedge$
- List operation: $+$



- <http://en.wikipedia.org/wiki/File:Venn0111.svg>

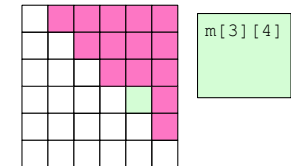
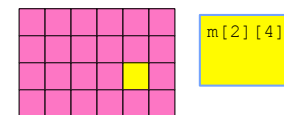
Patterns in some APTs

- **PositiveID and FriendScore APT (coming week)**

- Nested loops, look at everything: look at pairs
- While True: ... break

```
for i in range(4):
    for j in range(6):
        elt = m[i][j]

for i in range(6):
    for j in range(i+1, 6):
        elt = m[i][j]
```



How do we store/use tabular data?

- **How do we access data in FriendScore APT?**

- What does `friends[i][j]` mean?
 - What is `friends[i]`, what is `s[j]` where `s` is a string?

- **What is a matrix?**

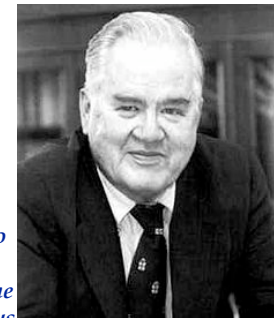
- Table of numbers? Entries? Two-Dimensional, $m \times n$
- Many arithmetic operations available
- Convenient for representing correlated data

- **In Python what is a list of lists?**

- Other languages have arrays, matrices, ...

John Tukey: 1915-2000

- Cooley-Tukey FFT
- Bit: Binary Digit
- Box-plot
- “software” used in print



Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.

The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.