

## Grammars, Regex, Problems and More

- **Grammars are used**
  - In computer science, designing software and hardware
  - In English, in Spanish, in all *natural* languages
  - In genomics, grammar of DNA?
- **Regular Expressions math, compsci, real problems**
  - How do recognize SPAM? Part statistics, part regex
  - How do we tell if email address entered is valid?
  - How do we search with wild-cards, e.g., `*@duke.edu`
- **How do we recognize valid Python program?**

Compsci 6/101, Spring 2012

14.1

## Grammars and Regex

```
<integer> ::= <digit> | <digit> <integer>
<digit> ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
```

- **Why is 1234 a valid integer? Is 01234 a valid integer?**
  - How could we avoid leading zeros?
  - What about a floating point number?
- **Regular expressions: mathematical and applied**
  - Create regexps from `.` `+` `*` `(` `|` `\` `$`
  - Understanding how these work best done by example
    - `[A-Za-z]+\.[A-Za-z]+@` and then more
  - <http://pdos.csail.mit.edu/scigen/>

Compsci 6/101, Spring 2012

14.2

## Regular Expressions

- **a, a+, a\*, [abc], [a-z], ^a, a\$**
  - These are useful in and of themselves, madlibs, RSG
  - Also good for exploring problems and grammars
- **Parsing and handling HTML**
  - Finding `<a href="http://...`
  - Why is this useful to Bing | Google?
- **Lots of details, more of a preview**
  - Where to look for more info?

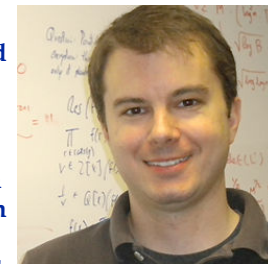


Compsci 6/101, Spring 2012

14.3

## Craig Gentry

Duke '95, Harvard Law, Stanford  
Compsci PhD  
ACM 2010 Hopper Award for...



"Fully homomorphic encryption is a bit like enabling a layperson to perform flawless neurosurgery while blindfolded, and without later remembering the episode. We believe this breakthrough will enable businesses to make more informed decisions, based on more studied analysis, without compromising privacy."

IBM VP, Software Research

Compsci 6/101, Spring 2012

14.4

## Compsci 6/101: Re[gex|cursion]

- **Recursion: self-referential structures and code**
  - Look up recursion in [Google | Bing]
  - Look-it up in the index of ...
  - What is a folder on your computer's desktop?
- **Powerful tool for elegantly expressing algorithms**
  - Never necessary, but alternative can be hard to develop, lengthy, tricky, ... (but then again ...)
  - Part of essential toolkit of computer scientist
    - Arguably not essential for web developer, entrepreneur, social media promoter, ...

Compsci 6/101, Spring 2012

14.5

## What's the deal with self-reference?

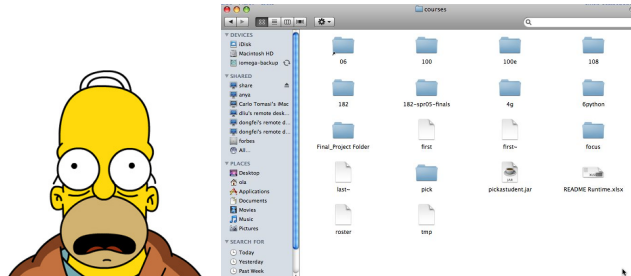
- ```
def visit(dirname):
    for file in dirname:
        if isdir(file): visit(file)
        else: print file
```
- **Does pseudo code make sense?**
    - Details make this a little harder in Python, but close!
  - **Recursive functions**
    - Function doesn't call itself, each function is a separate "thing", with its own state
    - Must have a *base case*, no recursive calls made, no self-referential work done

Compsci 6/101, Spring 2012

14.6

## What's in a folder on your computer?

- **Where are the large files? How do you find them?**
  - Can a folder be inside a folder? Why?



Compsci 6/101, Spring 2012

14.7

## Finding large files: FileVisit.py

```
def bigfiles(dirname,min_size):
    large = []
    #print dirname
    for sub in os.listdir(dirname):
        path = os.path.join(dirname,sub)
        if os.path.isdir(path):
            large.extend(bigfiles(path,min_size))
        else:
            size = os.path.getsize(path)
            if size > min_size:
                large.append((path,size))
    return large

big = bigfiles("c:\Users",10000)
[(file,102030),(nfile,1030303),(pfile,10001)]
```

Compsci 6/101, Spring 2012

14.8

## Dissecting FileVisit.py

- How do we find the contents of a folder?
  - Another name for folder: directory
  - How do we identify folder? (by name)
  - `os.listdir(dirname)` returns a list of ...
  - Path is `c:\user\ola\foo` or `/Users/ola/bar`
  - `os.path.join(dir,sub)` returns full path
  - Platform independent paths
- What's the difference between file and folder?
  - `os.path.isdir()` and `os.path.getsize()`

Compsci 6/101, Spring 2012

14.9

## Creativity with self-reference

- Sometimes madlibs are fun (corollary?)
  - Humans fill in the blanks
  - Computers automatically fill in the blanks

The <apt-name> APT was really <description> but I didn't do it because I <excuse>

<description> :: "cool", "terrible", "baller", ...  
<excuse> :: "was too tired", "didn't know how", ...  
<excuse> :: <excuse> and <excuse>

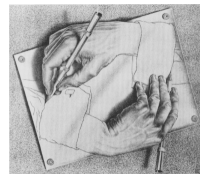
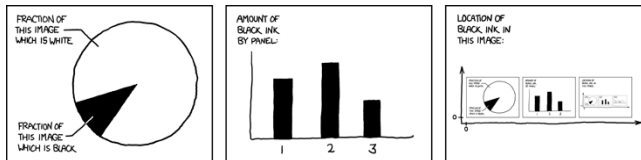
- See SimpleGrammar.py

Compsci 6/101, Spring 2012

14.10

## Recursion in Pictures

- <http://xkcd.com/688/> and <http://xkcd.com/543/>



Compsci 6/101, Spring 2012

14.11

## The power of regular expressions

- Interdisciplinary:
  - Music and Compsci (for Compsci 108 final project)
- Who is Ge Wang?
  - <http://www.youtube.com/watch?v=ADEFHmkL3HBg>

The final product is so much more than we had hoped for though it was something that we aimed for from the beginning.



Our investment into a huge and meticulous design process was a huge factor in making later progress. 35000+ lines of code / design / documentation gave us a project we were all very happy and proud to be a part of.

Compsci 6/101, Spring 2012

14.12

## Grammars for fun and recursion

- <http://en.wikipedia.org/wiki/SCIgen>
- <http://www.elsewhere.org/pomo/>
- <http://www-cs-faculty.stanford.edu/~zelenski/rsg/>
- I need an extension because <plea>.
- <plea>:
  - > <dubious-excuse>,
  - > <dubious-excuse> and <plea>

Compsci 6/101, Spring 2012

14.13

## Sierpinski Gasket

- [http://en.wikipedia.org/wiki/Sierpinski\\_triangle](http://en.wikipedia.org/wiki/Sierpinski_triangle)
- How to think of creating it?
  - > Remove central triangle, repeat recursively
  - > Make three half-sized triangles, join, repeat
  - > Chaos Game: completely at random!
- See links to L-system generation
  - > [http://en.wikipedia.org/wiki/L-system#Example\\_6:\\_Sierpinski\\_triangle](http://en.wikipedia.org/wiki/L-system#Example_6:_Sierpinski_triangle)
  - > Uses grammar!

Compsci 6/101, Spring 2012

14.14

## Koch Snowflake



Compsci 6/101, Spring 2012

14.15

## Sheryl Sandberg, COO Facebook

[http://www.ted.com/talks/sheryl\\_sandberg\\_why\\_we\\_have\\_too\\_few\\_women\\_leaders.html](http://www.ted.com/talks/sheryl_sandberg_why_we_have_too_few_women_leaders.html)

Sandberg says she eventually realized that women, unlike men, encountered tradeoffs between success and likability. The women had internalized self-doubt as a form of self-defense: people don't like women who boast about their achievements. The solution, she began to think, lay with the women. She blamed them more for their insecurities than she blamed men for their insensitivity or their sexism.



Compsci 6/101, Spring 2012

New Yorker, 7/11/2011<sup>14.16</sup>