

From grammars to regular expressions

- **Grammars are used**
 - In computer science
 - In English, in Spanish, in all *natural* languages
 - In genomics, grammar of DNA?
- **Parsing languages, what does parse mean?**
 - Computer and other, how do we process statements in an automatic way?
 - How do we generate statements that are "correct"?

Compsci 06/101, Spring 2011

15.1

Regular Expressions and Grammars

- **Grammar hierarchy in terms of 'power'**
 - Mathematical, formal model of language
 - Theoretically and practically interesting, e.g., processing spam, training spam filters, quickly looking through millions of files with *grep*
- <http://pdos.csail.mit.edu/scigen/>



Compsci 06/101, Spring 2011

15.2

Grammars and Regex

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

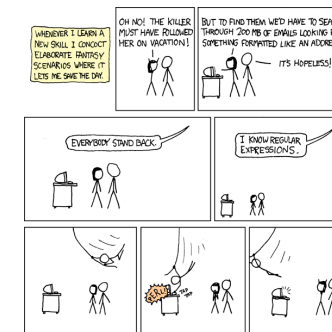
- **Why is 1234 a valid integer? Is 01234 a valid integer?**
 - How could we avoid leading zeros?
 - What about a floating point number?
- **Regular expressions: mathematical and applied**
 - Create regexps from . + * (| \ \$
 - Understanding how these work best done by example

Compsci 06/101, Spring 2011

15.3

What good are regular expressions

- <http://xkcd.com/208/>



Compsci 06/101, Spring 2011

15.4