38-year-old Prabhat Sinha carries a load of coal weighing 60kgs, supported by a head-strap, as he ascends the staircase of a coal mine on April 16, 2011 in Jaintia Hills, India. After traversing treacherous mountain roads, the coal is delivered to neighboring Bangladesh and to Assam from where it is distributed all over India, to be used primarily for power generation and as a source of fuel in cement plants. (Daniel Berehulak/Getty Images)

Files in Python

Lecture 10 - COMP 364

February 8, 2010, updated 2012

Mathieu Perreault

Combining Iteration and If-Else

```
core_list = []
periphery_list = []
for line in open('species.csv'):
  mylist = line.split(',')
  stype = mylist[1] # Can be core or periphery
  sname = mylist[2] # Name of species
  if stype == "core":
    core_list.append(sname)
  elif stype == "periphery":
    periphery_list.append(sname)

print "Core: ", len(core_list)
print "Periphery: ", len(periphery_list)
```

flow control

- one can use a **boolean** as a ‘switch’ to control the flow of a program.

```
foundit = False
for line in open(‘genes.fasta’):
  if ‘ABAT01000005.1’ in line:
    foundit = True
  else:
    if ‘ABAT01000006.1’ in line:
      foundit = False
    else:
      if foundit:
        print line
```

```
TAGCATACGACTGACGATG
>....ABAT01000005.1
GATCTAGACTACGCAGACT
GACGACTAGCACTACGACT
GAGCATATATCGGACTGAC
>....ABAT01000006.1
GACTAGCAGCATCACGACT
```

flow control

- one can use a **boolean** as a ‘switch’ to control the flow of a program.

```
foundit = False
for line in open(‘genes.fasta’):
  if ‘ABAT01000005.1’ in line:
    foundit = True
  else:
    if ‘ABAT01000006.1’ in line:
      foundit = False
    else:
      if foundit:
        print line
```

```
TAGCATACGACTGACGATG
>....ABAT01000005.1
GATCTAGACTACGCAGACT
GACGACTAGCACTACGACT
GAGCATATATCGGACTGAC
>....ABAT01000006.1
GACTAGCAGCATCACGACT
```

foundit
false->true
true
true->false

elif and flow control

- **elif** is a combination of else and if. Makes end result more readable.

```
foundit = False
for line in open(‘genes.fasta’):
  if ‘ABAT01000005.1’ in line:
    foundit = True
  else:
    if ‘ABAT01000006.1’ in line:
      foundit = False
    else:
      if foundit:
        print line
```

```
foundit = False
for line in open(‘genes.fasta’):
  if ‘ABAT01000005.1’ in line:
    foundit = True
  elif ‘ABAT01000006.1’ in line:
    foundit = False
  elif foundit:
    print line
```

File operations

- Reading content
- Writing content

- *The file workflow*
    - *Open the file*
    - *Read/write content*
    - *Close the file*

Reading Files - Technique 1

```
fh = open(<path to file>, ‘r’)

content = fh.readlines()

fh.close()

for line in content:

   <do stuff with line>
```

Exercise: Print lines that are less than 10 characters in length.

Reading Files - Technique 2

```
for line in open(‘/data/string/protein.links.fa’):

     <do stuff with the line>
```

This is the better technique for handling large files.

Writing content

```
fh = open(“/studentbox/output.txt”, ‘w’)

fh.write(‘Hello, world!’)

fh.close()
```

Exercise: Write all lines in *source.txt* that have less than 10 characters to file *output.txt*

Extra: string formatting

- When you want to print complex strings, use string substitution

- Main substitution types: %s (string), %d (digits/integers), %f (float)

```
print “This sequence is %d nucleotides long” % len(line)
print “For taxon ID %d, Species name is: %s” % (mylist[0],
 mylist[1])
```