Compsci 101
Stable Sorting

Susan Rodger
March 31, 2022

T is for …

• Type  
  • From int to float to string to list to …

• Text  
  • From .txt to editors to …

• Turing Award  
  • Nobel, Fields, Turing
  • Turing Duke Alums:  
    • Ed Clarke (MS)
    • John Cocke (BS, PhD)
    • Fred Brooks (BS)

Announcements

• APT-6 due Thursday, March 31, TODAY
• Assignment 5 Clever GuessWord due Tues. April 5
• APT-7 out today! Due April 7
• Lab 10 Friday  
  • There is a prelab!

• APT Quiz 2 in one week, April 7-10  
  • Covers topics thru APT 7
• Exam 4 is April 12, more on that on Tuesday
APT Quiz 2 April 7-10

- Opens 4/7 11:30am
- Closes at 11pm 4/10 – must finish all by this time
- There are two parts based on APTs 1-7
  - Each part has two APT problems
  - Each part is 1.5 hours – more if you get accommodations
  - Each part starts in Sakai under tests and quizzes
  - Sakai is a starting point with countdown timer that sends you to a new apt page just for each part
  - Could do each part on different day or same days
- Will put up problems today from an old APT Quiz so you can practice (not for credit) – on APT Page

APT Quiz 2

- Is your own work!
  - No collaboration with others!
  - Use your notes, lecture notes, your code, textbook
  - DO NOT search for answers!
  - Do not talk to others about the quiz until grades are posted
- Post private questions on Ed Discussion
  - We are not on between 10pm and 8am!
  - We are not on all the time
  - Will try to answer questions between 8am – 10pm
- See 101 APT page for tips on debugging APTs

PFTD

- Stable Sorting
- Some APTs

How is the sorting happening?

```python
>>> d
{'a': [1, 2, 3], 'b': [4, 7], 'c': [1, 1, 5, 8]}
>>> sorted(d.items())

>>> sorted(d.items(), key=lambda x: x[1])

>>> sorted(d.items(), key=lambda x: x[1][-1])
```
How to do some “fancy” sorting

- lambda PARAMETER: EXPRESSION

- Given data: list of tuples: (first name, last name, age)
  
  ```
  [('Percival', 'Avram', 51),
   ('Melete', 'Sandip', 24), ...]
  ```

- Think: What is the lambda key to sort the following?

  ```
  sorted(data, key=lambda z : (z[0],z[1],z[2]))
  ```

  - Sort by last name, break ties with first name
  - Sort by last name, break ties with age
  - Alphabetical by last name, then first name, then reverse age order
Creating Tuples with lambda

- Sort by last name, break ties with first name
- Sort by last name, break ties with age
- Alphabetical by last name, then first name, then reverse age order

Leveraging the Algorithm

- Can’t sort by creating a tuple with lambda, use:
  - Pattern: Multiple-pass *stable* sort – first sort with last tie breaker, then next to last tie breaker, etc. until at main criteria
  - Sort by index 0, break tie in reverse order with index 1
    
    [('b', 'z'), ('c', 'x'), ('b', 'x'), ('a', 'z')]
    
    - *Stable* sort respects original order of "equal" keys

Creating Tuples with lambda

- Sort by last name, break ties with first name
  - key = lambda x: (x[1], x[0])
- Sort by last name, break ties with age
  - key = lambda x: (x[1], x[2])
- Alphabetical by last name, then first name, then reverse age order
  - key = lambda x: (x[1], x[0], -x[2])
- What if wanted something really different?
  - Sort alphabetical by last name, break ties by reverse alphabetical using first name

Leveraging the Algorithm

- Can’t sort by creating a tuple with lambda, use:
  - Pattern: Multiple-pass *stable* sort – first sort with last tie breaker, then next to last tie breaker, etc. until at main criteria
  - Sort by index 0, break tie in reverse order with index 1
    
    [('b', 'z'), ('c', 'x'), ('b', 'x'), ('a', 'z')]
    
    [('b', 'z'), ('a', 'z'), ('c', 'x'), ('b', 'x')]
    
    [('a', 'z'), ('b', 'z'), ('b', 'x'), ('c', 'x')]
    
    - *Stable* sort respects original order of "equal" keys
Leveraging the Algorithm

- Can't sort by creating a tuple with lambda, use:
  - Pattern: Multiple-pass **stable** sort – first sort with last tie breaker, then next to last tie breaker, etc. until at main criteria

- **Sort by index 0**, break tie in reverse order with index 1
  
  
  
  
  
  

- **Stable** sort respects original order of "equal" keys

Stable sorting: respect "equal" items

- Women before men, each group height-sorted
  - First sort by height

- Then sort by gender

Understanding Multiple-Pass Sorting

```python
> data
[(‘f’, 2, 0), (‘e’, 1, 4), (‘a’, 2, 0), (‘c’, 2, 5), (‘b’, 3, 0), (‘d’, 2, 4)]
> a0 = sorted(data, key = lambda x: x[0])
a0

> a1 = sorted(a0, key = lambda x: x[2])
a1

> a2 = sorted(a1, key = lambda x: x[1])
a2
```
### Understanding Multiple-Pass Sorting

```python
> data
[(f', 2, 0), (e', 1, 4), (a', 2, 0),
(c', 2, 5), (b', 3, 0), (d', 2, 4)]
> a0 = sorted(data, key = lambda x: x[0])
> a0
[(a', 2, 0), (b', 3, 0), (c', 2, 5),
(d', 2, 4), (e', 1, 4), (f', 2, 0)]
> a1 = sorted(a0, key = lambda x: x[2])
> a1
[(a', 2, 0), (b', 3, 0), (f', 2, 0),
(d', 2, 4), (e', 1, 4), (c', 2, 5)]
> a2 = sorted(a1, key = lambda x: x[1])
> a2
[(e', 1, 4), (a', 2, 0), (f', 2, 0),
(d', 2, 4), (c', 2, 5), (b', 3, 0)]
```

---

### WOTO-1 Multi-pass Sorting

Unpack from Inside out

```python
lst = [[4, 6, 7], [5, 2], [3, 9], [6, 2, 9]]
x = sorted(lst, key=sum)
y = sorted(x, key=min)
z = sorted(y, key=max)
```

Unpack from Inside out

```python
lst = [[4, 6, 7], [5, 2], [3, 9], [6, 2, 9]]
x = sorted(lst, key=sum)
y = sorted(x, key=min)
z = sorted(y, key=max)
```

When and What’s in CompSci 101

- Problem to solve
  - Use 7 steps
  - Step 5: How do you translate algorithm to code?
    - What do you use to solve it?
    - When do you use it?

What are the “what’s”?

- Data Structures: list, set, dictionary, tuple
- Loops: from for to while
- Other:
  - List comprehensions
  - Parallel lists
  - Lambda
  - If…if…if
  - If…elif…else
Quick When’s and What’s for 101

- Whichever makes more sense to you:
  - Parallel lists vs dictionaries
  - If...if...if vs if...elif...else
  - List comprehension vs for loop
  - Tuples vs Lists
    - If you want to prevent mutation -> tuples
  - Need single line function
    - Lambda vs create normal helper function

What's the best way to ...

- SortedFreqs
  - [https://www2.cs.duke.edu/csed/pythonapt/sortedfreqs.html](https://www2.cs.duke.edu/csed/pythonapt/sortedfreqs.html)

- Count how many times each string occurs
  - Create d = {}, iterate over list updating values

- OR
  - Use data.count(w) for each w

APT – Sorted Freqs

**Problem Statement**

The frequency with which data occurs is sometimes an important statistic. In this problem you'll determine how frequently strings occur and return a list representing the frequencies of each different/unique string. The list returned contains as many frequencies as there are unique strings. The returned frequencies represent an alphabetic/lexicographic ordering of the unique words, so the first frequency is how many times the alphabetically first word occurs and the last frequency is the number of times the alphabetically last word occurs.

Consider these strings (quotes for clarity, they're not part of the strings).

```python
"apple", "pear", "cherry", "apple", "cherry", "pear", "apple", "banana"
```

The list returned is (1, 1, 2, 2) since the alphabetically first word is "apple" which occurs 3 times; the second word alphabetically is "banana" which occurs once, and the other words each occur twice.

**Specification**

```python
def freqs(data):
    ***
    return list of int values corresponding to frequencies of strings in data, a list of strings
    ***
```

```python
filename: SortedFreqs.py
def freqs(data):
    ***
    return list of int values corresponding to frequencies of strings in data, a list of strings
    ***
```

What's the best way to ...

- SortedFreqs
  - [https://www2.cs.duke.edu/csed/pythonapt/sortedfreqs.html](https://www2.cs.duke.edu/csed/pythonapt/sortedfreqs.html)

- Count how many times each string occurs
  - Create d = {}, iterate over list updating values

- OR
  - Use data.count(w) for each w

- Wait, that looks like …

```python
def freqs(data):
    return [data.count(d) for d in sorted(set(data))]
```
APT: SortByFreqs

Problem Statement
The frequency with which data occurs is sometimes an important statistic. In this problem you are given a list of strings and must determine how frequently the strings occur. Return a list of strings that is sorted (ordered) by frequency. The first element of the returned list is the most frequently occurring string, the last element is the least frequently occurring. Ties are broken by listing strings in lexicographic/alphabetical order. The returned list contains one occurrence of each unique string from the list parameter.

Consider these strings (quotes for clarity, they’re not part of the strings): 

"apple", "pear", "cherry", "apple", "pear", "apple", "banana"

The list returned is:

[ "apple", "pear", "banana", "cherry" ]

since the most frequently occurring string is "apple" which occurs 3 times, the string "pear" occurs twice and the other strings each occur once so they are returned in alphabetical order.

SortByFreqs Example

```python
data = ["apple", "pear", "cherry", "apple", "pear", "apple", "banana"]
tups = [(3, "apple"),(1, "banana"),(1,"cherry"),(2,"pear")]

sorted(...) line 8 = [(3,"apple"),(2,"pear"),(1, "banana"),(1,"cherry")]

result = ["apple","pear","banana","cherry"]
```

Wait, wait, but what's ...

- SortByFreqs
  - [https://www2.cs.duke.edu/csed/pythonapt/sortbyfreqs.html](https://www2.cs.duke.edu/csed/pythonapt/sortbyfreqs.html)

- Sort by # occurrences high to low
  - Tuples with count/lambda and reverse=True?
  - Break ties in alphabetical order: two passes

```
6  def sort(data):
7      tups = [(data.count(t),t) for t in sorted(set(data))]
8      result = [t[1] for t in sorted(tups,key=lambda x : x[0],reverse=True)]
9                      return result
```
SortByFreqs Example

- SortByFreqs
  - [https://www2.cs.duke.edu/csed/pythonapt/sortbyfreqs.html](https://www2.cs.duke.edu/csed/pythonapt/sortbyfreqs.html)

```python
def sort(data):
    tups = [(data.count(t), t) for t in sorted(set(data))]
    result = [t[1] for t in sorted(tups, key=lambda x: x[0], reverse=True)]
    return result

data = ['apple', 'pear', 'cherry', 'apple', 'pear', 'apple', 'banana']
tups = [(3, 'apple'), (1, 'banana')]
print(sorted(...) line 8 = [(3, 'apple'), (1, 'banana')]
result = ['apple', 'pear', 'banana', 'cherry']
```

PRINT A LOT!

APT: MedalTable

Tracking the Data

- What do we need to obtain for each country?
  - What's the data, how do we store it?
  - What's the data, how do we calculate it?

- Method and code to transform input
  - What will we store, how do we initialize/update
  - Verifying we've done this properly

Use a dictionary?

- 3 dictionaries
  - Country to Gold count
  - Country to Silver count
  - Country to Bronze count

Example

- How would we create a dictionary for:

  ```
  ["KOR TPE UKR", "KOR KOR TPE", "KOR JPN JPN"]
  ```

- 1 dictionary
  - Map country to [gld cnt, sil cnt, bro cnt]
Example: dictionary d

- Process first string: "KOR TPE UKR"
  - \(d["KOR"] = [1, 0, 0]\) # gold
  - \(d["TPE"] = [0, 1, 0]\) # silver
  - \(d["UKR"] = [0, 0, 1]\) # bronze

- Process second string: "KOR KOR TPE"
  - \(d["KOR"] = [2, 0, 0]\) # gold
  - \(d["KOR"] = [2, 1, 0]\) # silver
  - \(d["TPE"] = [0, 1, 1]\) # bronze

Example: dictionary d (2)

- What we have so far:
  - \(d["KOR"] = [2, 1, 0]\)
  - \(d["TPE"] = [0, 1, 1]\)
  - \(d["UKR"] = [0, 0, 1]\)

- Process third string: "KOR JPN JPN"
  - \(d["KOR"] = [3, 1, 0]\) # gold
  - \(d["JPN"] = [0, 1, 0]\) # silver
  - \(d["JPN"] = [0, 1, 1]\) # bronze
Sorting the Data

- Write a helper function to build the dictionary
  - \( d = \text{builddict} \text{(results)} \)
  - Where results is string of countries for each event
- Use dictionary to get list of tuples
  
  \[
  [(\text{'JPN'}, [0, 1, 1]), (\text{'KOR'}, [3, 1, 0]), (\text{'TPE'}, [0, 1, 1]), (\text{'UKR'}, [0, 0, 1])]
  \]

- Then do passes to sort the data
  - Will discuss sorting the data in lab

Some of the code

```python
def builddict(data):
    d = {}
    for item in data:
        [gld, sil, bro] = item.split()
```

Use three variables for the three countries

```python
if gld not in d:
    d[gld] = [0, 0, 0]
```

If country not in, set to \([0, 0, 0]\) since has no medals counted yet
Some of the code

```python
def build_dict(data):
    d = {}  
    for item in data:
        [gld, sil, bro] = item.split()
        if gld not in d:
            d[gld] = [0, 0, 0]
        d[gld][0] += 1
```

Update appropriate index, [0] is the gold count