

Searching

Objectives

- Students will demonstrate how combining data sources and classifying data are part processing data
- Students will describe challenges to structuring large data sets for analysis
- Students will identify how the order of data influences which methods are appropriate for searching the data.
- Students will describe standard search algorithms in pseudocode and in Python.
- Students will Compare different algorithms for *efficiency* when searching for an item.

What would be the best way to organize books on a family bookshelf so that you could find the one you wanted very quickly? Would you need a different method for different sizes of collections?

Methods of Searching



Linear Search Activity

Checking every single thing in
the order it is in



Binary Search Contest

Sorted!



Binary Search Activity

Divide in half
check higher
lower or
match
just new "half"

Reminder of how to use files

variable

```
fout = open(filename, "r")
```

variable

```
for i in fout
```

The diagram illustrates the flow of data in a Python script. At the top, the word "variable" is written in blue. A blue arrow points from this word down to the parameter "filename" in the code snippet "fout = open(filename, 'r')". Below this, another blue arrow points from the variable "fout" in the same code snippet down to the variable "fout" in the code snippet "for i in fout". A third blue arrow points from the word "variable" written below "fout" in the first snippet up to the "fout" variable in the same snippet.

number in the list
number to search for

Google
classroom

Coding a Linear Search

boolean variable = false

for loop runs from start to end

if number in list = search value

we found it
break out of loop

boolean = true

else

repeat for next

Pseudocode

if boolean = false
not in the list

Coding a Binary Search

* has to be sorted

length = len(list name)

start = 0

mid = (end - start) / 2

end = length

while (start ≤ end)

if (my Number = list[mid])
found

elif (my Number < list)

end = mid

mid =

else

start = mid

mid =

Search Options Comparison

1. Use your Sequential (Linear) Search programs with the given datasets to fill out the following table:

		Linear Search		Binary Search	
DataSet	Number	Found/ Not Found	Number of Items Checked	Found/ Not Found	Number of Items Checked
DataRand100.csv	77				
DataRand1000.csv	780				
DataRand10000.csv	2735				
DataSorted100.csv	138				
DataSorted1000.csv	875				
DataSorted10000.csv	54798				