Objectives

- More on using strings
- Computer's representations of data types
- **str** methods

Strings

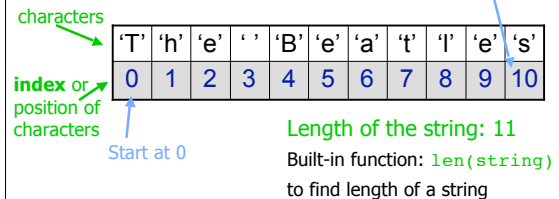- Actually a *sequence* of characters
  - Example:
    string = "The Beatles"

characters

End at len() -1

'T'	'h'	'e'	' '	'B'	'e'	'a'	't'	'l'	'e'	's'
0	1	2	3	4	5	6	7	8	9	10

index or position of characters

Start at 0

Length of the string: 11
Built-in function: `len(string)`
to find length of a string

Iterating Through a String

- For each character in the string

string of length 1

```
for char in string:
    print char
```

'T'	'h'	'e'	' '	'B'	'e'	'a'	't'	'l'	'e'	's'
0	1	2	3	4	5	6	7	8	9	10

An integer

```
for pos in xrange(len(string)):
    print string[pos]
```

Index into the string

- For each position in the string

Substrings Operator []

- Look at a particular character in the string
  - Syntax: str[<integer expression>]
  - [Positive values]: index of character
  - [Negative values]: count backwards from end
- Look at a sequence of characters in the string
  - Syntax: str[<start>:<end>]

Testing for Substrings

- Using the **in** operator
  - Used **in** before in **for** loops
- Syntax:
    substring **in** string
  - Evaluates to True or False
- Example:

```
if "cat" in name:
    print name, "contains 'cat'"
```

Representations of Data

- Computer needs ways to represent different types of data
  - Eventually, all boils down to 1s and 0s
- Computer needs to translate between what we know (**decimal, strings**) to what the computer knows (**binary**) and back again

- Seems like a divergence on strings but just wait…

Decimal Representations

- Decimal is base 10
- Digits: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9
- Each *position* in a decimal number represents a power of 10

Decimal Representations

- Decimal is base 10
- Digits: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9
- Each *position* in a decimal number represents a power of 10
- Example: 54,087

5	4	0	8	7
10^4	10^3	10^2	10^1	10^0

- $= 5*10^4 + 4*10^3 + 0*10^2 + 8*10^1 + 7*10^0$
- $= 5*10,000 + 4*1000 + 0*100 + 8*10 + 7*1$

Number Representations

Characteristic	Decimal	Binary
Base	10	2
Digits	0, 1, 2, 3, 4, 5, 6, 7, 8, 9	0, 1
Position represents	Power of 10	Power of 2

- Binary: two values (0, 1)
  - ➤ Like a light switch (either **off** or **on**) or booleans (either True or False)
- 0 and 1 are *binary digits* or **bits**
  - ➤ 64-bit machine: represents numbers (and other data) with 64 bits

Binary Representation

- Example: 1101

1	1	0	1
2^3	2^2	2^1	2^0

- $= 1*2^3 + 1*2^2 + 0*2^1 + 1*2^0$
- $= 1*8 + 1*4 + 0*2 + 1*1$
  - ➤ Decimal value: 13

- **Practice**: 10110

Binary Representation

- Example: 10110

1	0	1	1	0
2^4	2^3	2^2	2^1	2^0

- $= 1*2^4 + 0*2^3 + 1*2^2 + 1*2^1 + 0*2^0$
- $= 1*16 + 0*8 + 1*4 + 1*2 + 0*1$
  - ➤ 22

Generalize this process into an algorithm...

Algorithm: Converting Binary to Decimal

- Accumulator design pattern
- Read in the binary number as a string
  - ➤ The starting exponent will be the length of the string-1
- Initialize the result to zero
- For each bit in the binary number
  - ➤ Multiply the bit by the appropriate power of 2
  - ➤ Add this to the result
  - ➤ Reduce the exponent by 1
- Print the result

Implement algorithm binaryToDecimal.py

Algorithm: Converting Decimal to Binary

- Read in the decimal as an integer
- Initialize the result to the empty string
- Repeat until the decimal is 0:
  - result = str(decimal % 2) + result
  - decimal = decimal / 2
- Print the result

Try out algorithm with 22

Implement algorithm
decimalToBinary.py

Feb 13, 2008   Sprenkle - CS111   13

String Representations

- A **string** is a sequence of characters
- Each character is stored as a binary number
- **ASCII** (American Standard Code for Information Interchange) is one standard encoding for characters
  - One problem with ASCII is that it is based on the English language
  - Cannot represent other types of characters
- Unicode is a new standard

Feb 13, 2008   Sprenkle - CS111   ASCII Table Handout   14

ASCII Questions

- Lowercase letters are represented by what range of numbers?

- Uppercase letters are represented by what range of numbers?

- What is the difference between the decimal encoding of 'M' and 'N'?
  - Between 'm' and 'n'?

Feb 13, 2008   Sprenkle - CS111   15

ASCII Questions

- Lowercase letters are represented by what range of numbers?
  - 97--122
- Uppercase letters are represented by what range of numbers?
  - 65--90
- What is the difference between the decimal encoding of 'M' and 'N'?
  - Between 'm' and 'n'?
  - 1

Feb 13, 2008   Sprenkle - CS111   16

Translating to/from ASCII

- Translate a character into its ASCII numeric code using **ord**
  - ord('a') ==> 97
- Translate an ASCII numeric code into its character using **chr**
  - chr(97) ==> 'a'

ascii_table.py
Feb 13, 2008   Sprenkle - CS111   ascii.py   17

Encryption

- Process of encoding information to keep it secure
- One technique: Substitution Cipher
  - Each character in message is replaced by a new character

Feb 13, 2008   Sprenkle - CS111   18

Caesar Cipher

- Replace with a character X places away
  - X is your *key*
- Julius Caesar used it to communicate with his generals
- "Wrap around"
- Write program(s) to do this in lab

Caesar Cipher

- Using the ASCII handout, what would be the encoded messages?

Message	Key	Encoded Message
apple	5	
zebra	5	
the eagle flies at midnight	-5	

Caesar Cipher

- What is your algorithm for the encoding process?
- How would you decode an encrypted message?

Message	Key	Encoded Message
apple	5	fuuqj
zebra	5	ejgwf
the eagle flies at midnight	-5	ocz zvbgz agdzn vo hdyidbco

Broader Issue: Electronic Voting

- Select from one of three articles
  - The Risks of Electronic Voting
  - Princeton Scientists Create Vote-Stealing Program
  - Voting with (Little) Confidence