Joyce Woznica  
IST 652, Week 9

Activity 9.3: Regular Expressions

﻿# Consider this example text, consisting of a Tweet, an excerpt from an HTML page,

# and an extra sentence.

text = '''Celebrate #NationalPetDay with our puppy playlist: https://t.co/eBHHFPW0z7 https://t.co/uix5AY2FFQ<a href="http://msande.stanford.edu"> Management Science and Engineering </a><p class="MsoNormal"

Address: Terman 311, Stanford CA 94305<br>

Email: ashishg@cs.stanford.edu<br>

Phone: (650)814-9999 [Cell], Fax: (650)723-9999<br>

Admin asst: Roz Morf, Terman 405, 650-723-9999, rozm@stanford.edu</p>

The U.S.A. olympic teams have east-west training centers with up-to-date equipment.

'''

# Write regular expressions RegExPal.com to find the following items,

# where these can all be separate regular expressions:

#

# Match the URLs (there should be three results).

https?:\/\/(www\.)?[-a-zA-Z0-9@:%.\_\+~#=]{1,256}\.[a-zA-Z0-9()]{1,6}\b([-a-zA-Z0-9()@:%\_\+.~#?&//=]\*)

# Match the phone numbers (three results).

(\([0-9]{3}\)[\s-]?|[0-9]{3}-)[0-9]{3}-[0-9]{4}

# Match the e-mail addresses (two results).

(?:[a-z0-9!#$%&'\*+/=?^\_`{|}~-]+(?:\.[a-z0-9!#$%&'\*+/=?^\_`{|}~-]+)\*|"(?:[\x01-\x08\x0b\x0c\x0e-\x1f\x21\x23-\x5b\x5d-\x7f]|\\[\x01-\x09\x0b\x0c\x0e-\x7f])\*")@(?:(?:[a-z0-9](?:[a-z0-9-]\*[a-z0-9])?\.)+[a-z0-9](?:[a-z0-9-]\*[a-z0-9])?|\[(?:(?:25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)\.){3}(?:25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?|[a-z0-9-]\*[a-z0-9]:(?:[\x01-\x08\x0b\x0c\x0e-\x1f\x21-\x5a\x53-\x7f]|\\[\x01-\x09\x0b\x0c\x0e-\x7f])+)\])

# Match acronyms (there should be two results, with or without dots: CA and U.S.A.).

#

# Use regexpal.com to text your expressions.

# Capture the expression that gives you the desired result and place in a single document.