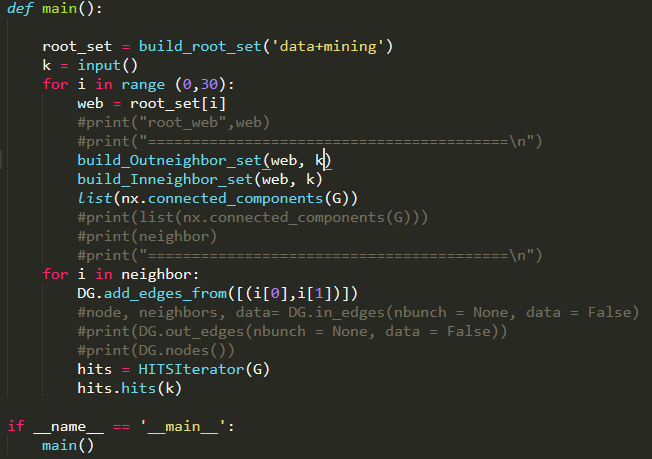
**Data\_mining advance project**

**Name: Shang-Yung Hsu**

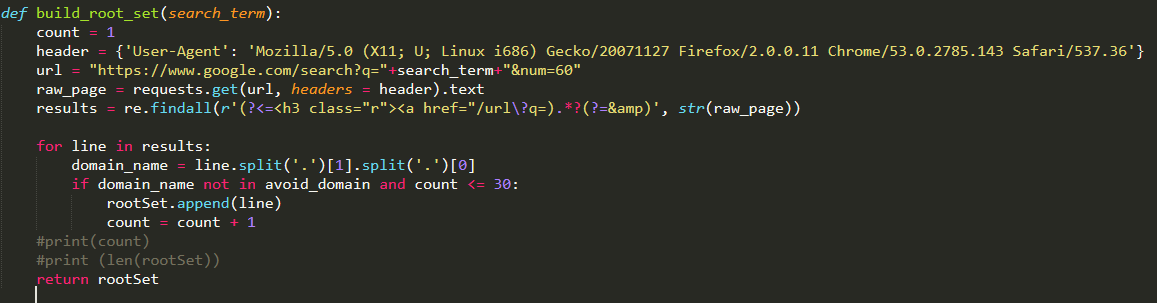
**ID: sh555**

Step:

Main:



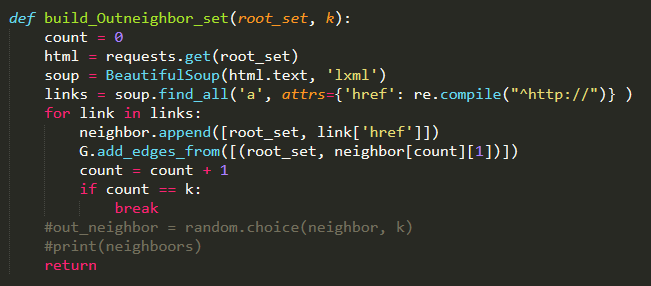
Finding the 30 root sets from the Google:



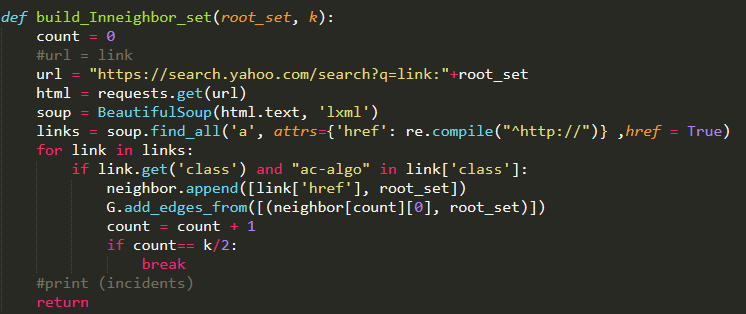
Also avoid the domain name those are



To build the neighbor which the root link to:



Build the neighbor which link to root:



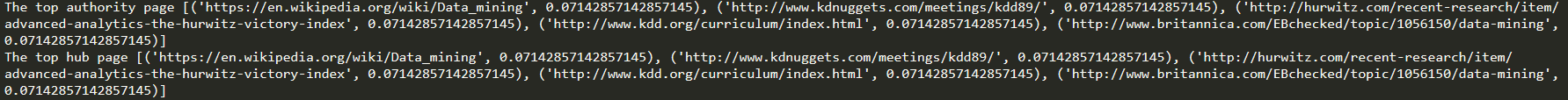
Run the Hits algo:



Result:

**For data mining**

**K = 5**



The top authority page [('https://en.wikipedia.org/wiki/Data\_mining', 0.07142857142857145), ('http://www.kdnuggets.com/meetings/kdd89/', 0.07142857142857145), ('http://hurwitz.com/recent-research/item/advanced-analytics-the-hurwitz-victory-index', 0.07142857142857145), ('http://www.kdd.org/curriculum/index.html', 0.07142857142857145), ('http://www.britannica.com/EBchecked/topic/1056150/data-mining', 0.07142857142857145)]

The top hub page [('https://en.wikipedia.org/wiki/Data\_mining', 0.07142857142857145), ('http://www.kdnuggets.com/meetings/kdd89/', 0.07142857142857145), ('http://hurwitz.com/recent-research/item/advanced-analytics-the-hurwitz-victory-index', 0.07142857142857145), ('http://www.kdd.org/curriculum/index.html', 0.07142857142857145), ('http://www.britannica.com/EBchecked/topic/1056150/data-mining', 0.07142857142857145)]

[Finished in 34.3s]

**For k = 10**

The top authority page [('https://en.wikipedia.org/wiki/Data\_mining', 0.06819943394704736), ('http://www.kdnuggets.com/meetings/kdd89/', 0.06819943394704736), ('http://hurwitz.com/recent-research/item/advanced-analytics-the-hurwitz-victory-index', 0.06819943394704736), ('http://www.kdd.org/curriculum/index.html', 0.06819943394704736), ('http://www.britannica.com/EBchecked/topic/1056150/data-mining', 0.06819943394704736), ('http://www-stat.stanford.edu/~tibs/ElemStatLearn/', 0.06819943394704736), ('http://proquest.safaribooksonline.com.proxy.library.carleton.ca/book/databases/data-warehouses/9780123814791', 0.06819943394704736), ('http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf', 0.06819943394704736), ('http://www.okairp.org/documents/2005%20Fall/F05\_ROMEDataQualityETC.pdf', 0.06819943394704736), ('http://www.kdnuggets.com/data\_mining\_course/x1-intro-to-data-mining-notes.html', 0.06819943394704736)]

The top hub page [('https://en.wikipedia.org/wiki/Data\_mining', 0.06819943394704736), ('http://www.kdnuggets.com/meetings/kdd89/', 0.06819943394704736), ('http://hurwitz.com/recent-research/item/advanced-analytics-the-hurwitz-victory-index', 0.06819943394704736), ('http://www.kdd.org/curriculum/index.html', 0.06819943394704736), ('http://www.britannica.com/EBchecked/topic/1056150/data-mining', 0.06819943394704736), ('http://www-stat.stanford.edu/~tibs/ElemStatLearn/', 0.06819943394704736), ('http://proquest.safaribooksonline.com.proxy.library.carleton.ca/book/databases/data-warehouses/9780123814791', 0.06819943394704736), ('http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf', 0.06819943394704736), ('http://www.okairp.org/documents/2005%20Fall/F05\_ROMEDataQualityETC.pdf', 0.06819943394704736), ('http://www.kdnuggets.com/data\_mining\_course/x1-intro-to-data-mining-notes.html', 0.06819943394704736)]

**For k = 20**

The top authority page [('https://en.wikipedia.org/wiki/Data\_mining', 0.05933908290969247), ('http://www.kdnuggets.com/meetings/kdd89/', 0.05933908290969247), ('http://hurwitz.com/recent-research/item/advanced-analytics-the-hurwitz-victory-index', 0.05933908290969247), ('http://www.kdd.org/curriculum/index.html', 0.05933908290969247), ('http://www.britannica.com/EBchecked/topic/1056150/data-mining', 0.05933908290969247), ('http://www-stat.stanford.edu/~tibs/ElemStatLearn/', 0.05933908290969247), ('http://proquest.safaribooksonline.com.proxy.library.carleton.ca/book/databases/data-warehouses/9780123814791', 0.05933908290969247), ('http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf', 0.05933908290969247), ('http://www.okairp.org/documents/2005%20Fall/F05\_ROMEDataQualityETC.pdf', 0.05933908290969247), ('http://www.kdnuggets.com/data\_mining\_course/x1-intro-to-data-mining-notes.html', 0.05933908290969247), ('http://www.kdd.org/explorations/view/june-1999-volume-1-issue-1', 0.05933908290969247), ('http://www.kdnuggets.com/polls/2002/methodology.htm', 0.05933908290969247), ('http://www.kdnuggets.com/polls/2004/data\_mining\_methodology.htm', 0.05933908290969247), ('http://www.kdnuggets.com/polls/2007/data\_mining\_methodology.htm', 0.05933908290969247), ('http://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html', 0.05933908290969247), ('http://cdn.intechopen.com/pdfs/5937/InTech-A\_data\_mining\_amp\_knowledge\_discovery\_process\_model.pdf', 0.05933908290969247), ('http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=451120', 0.05933908290969247), ('http://www.iadis.net/dl/final\_uploads/200812P033.pdf', 0.05933908290969247), ('http://academic.research.microsoft.com/?SearchDomain=2&SubDomain=7&entitytype=2', 0.05933908290969247), ('http://www.kdd.org/conferences.php', 0.05933908290969247)]

The top hub page [('https://en.wikipedia.org/wiki/Data\_mining', 0.05933908290969264), ('http://www.kdnuggets.com/meetings/kdd89/', 0.05933908290969264), ('http://hurwitz.com/recent-research/item/advanced-analytics-the-hurwitz-victory-index', 0.05933908290969264), ('http://www.kdd.org/curriculum/index.html', 0.05933908290969264), ('http://www.britannica.com/EBchecked/topic/1056150/data-mining', 0.05933908290969264), ('http://www-stat.stanford.edu/~tibs/ElemStatLearn/', 0.05933908290969264), ('http://proquest.safaribooksonline.com.proxy.library.carleton.ca/book/databases/data-warehouses/9780123814791', 0.05933908290969264), ('http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf', 0.05933908290969264), ('http://www.okairp.org/documents/2005%20Fall/F05\_ROMEDataQualityETC.pdf', 0.05933908290969264), ('http://www.kdnuggets.com/data\_mining\_course/x1-intro-to-data-mining-notes.html', 0.05933908290969264), ('http://www.kdd.org/explorations/view/june-1999-volume-1-issue-1', 0.05933908290969264), ('http://www.kdnuggets.com/polls/2002/methodology.htm', 0.05933908290969264), ('http://www.kdnuggets.com/polls/2004/data\_mining\_methodology.htm', 0.05933908290969264), ('http://www.kdnuggets.com/polls/2007/data\_mining\_methodology.htm', 0.05933908290969264), ('http://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html', 0.05933908290969264), ('http://cdn.intechopen.com/pdfs/5937/InTech-A\_data\_mining\_amp\_knowledge\_discovery\_process\_model.pdf', 0.05933908290969264), ('http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=451120', 0.05933908290969264), ('http://www.iadis.net/dl/final\_uploads/200812P033.pdf', 0.05933908290969264), ('http://academic.research.microsoft.com/?SearchDomain=2&SubDomain=7&entitytype=2', 0.05933908290969264), ('http://www.kdd.org/conferences.php', 0.05933908290969264)]

**For Deep Learning:**

**K = 5**

The top authority page [('https://en.wikipedia.org/wiki/Deep\_learning', 0.06108472217815262), ('http://adsabs.harvard.edu/abs/2015Natur.521..436L', 0.06108472217815262), ('http://ieeexplore.ieee.org/document/6248110/', 0.06108472217815262), ('http://adsabs.harvard.edu/abs/1996Natur.381..607O', 0.06108472217815262), ('http://research.microsoft.com/pubs/209355/DeepLearning-NowPublishing-Vol7-SIG-039.pdf', 0.06108472217815262)]

The top hub page [('https://en.wikipedia.org/wiki/Deep\_learning', 0.06108472217815262), ('http://adsabs.harvard.edu/abs/2015Natur.521..436L', 0.06108472217815262), ('http://ieeexplore.ieee.org/document/6248110/', 0.06108472217815262), ('http://adsabs.harvard.edu/abs/1996Natur.381..607O', 0.06108472217815262), ('http://research.microsoft.com/pubs/209355/DeepLearning-NowPublishing-Vol7-SIG-039.pdf', 0.06108472217815262)]

**K = 10**

The top authority page [('https://en.wikipedia.org/wiki/Deep\_learning', 0.06651901052377392), ('http://adsabs.harvard.edu/abs/2015Natur.521..436L', 0.06651901052377392), ('http://ieeexplore.ieee.org/document/6248110/', 0.06651901052377392), ('http://adsabs.harvard.edu/abs/1996Natur.381..607O', 0.06651901052377392), ('http://research.microsoft.com/pubs/209355/DeepLearning-NowPublishing-Vol7-SIG-039.pdf', 0.06651901052377392), ('http://sanghv.com/download/soft/machine%20learning,%20artificial%20intelligence,%20mathematics%20ebooks/ML/learning%20deep%20architectures%20for%20AI%20%282009%29.pdf', 0.06651901052377392), ('http://www.scholarpedia.org/article/Deep\_Learning', 0.06651901052377392), ('http://adsabs.harvard.edu/abs/2009SchpJ...4.5947H', 0.06651901052377392), ('http://deeplearning.cs.cmu.edu/pdfs/Cybenko.pdf', 0.06651901052377392), ('http://papers.nips.cc/paper/7203-the-expressive-power-of-neural-networks-a-view-from-the-width', 0.06651901052377392)]

The top hub page [('https://en.wikipedia.org/wiki/Deep\_learning', 0.06651901052377392), ('http://adsabs.harvard.edu/abs/2015Natur.521..436L', 0.06651901052377392), ('http://ieeexplore.ieee.org/document/6248110/', 0.06651901052377392), ('http://adsabs.harvard.edu/abs/1996Natur.381..607O', 0.06651901052377392), ('http://research.microsoft.com/pubs/209355/DeepLearning-NowPublishing-Vol7-SIG-039.pdf', 0.06651901052377392), ('http://sanghv.com/download/soft/machine%20learning,%20artificial%20intelligence,%20mathematics%20ebooks/ML/learning%20deep%20architectures%20for%20AI%20%282009%29.pdf', 0.06651901052377392), ('http://www.scholarpedia.org/article/Deep\_Learning', 0.06651901052377392), ('http://adsabs.harvard.edu/abs/2009SchpJ...4.5947H', 0.06651901052377392), ('http://deeplearning.cs.cmu.edu/pdfs/Cybenko.pdf', 0.06651901052377392), ('http://papers.nips.cc/paper/7203-the-expressive-power-of-neural-networks-a-view-from-the-width', 0.06651901052377392)]

**K = 20**

The top authority page [('https://en.wikipedia.org/wiki/Deep\_learning', 0.052999894000318216), ('http://adsabs.harvard.edu/abs/2015Natur.521..436L', 0.052999894000318216), ('http://ieeexplore.ieee.org/document/6248110/', 0.052999894000318216), ('http://adsabs.harvard.edu/abs/1996Natur.381..607O', 0.052999894000318216), ('http://research.microsoft.com/pubs/209355/DeepLearning-NowPublishing-Vol7-SIG-039.pdf', 0.052999894000318216), ('http://sanghv.com/download/soft/machine%20learning,%20artificial%20intelligence,%20mathematics%20ebooks/ML/learning%20deep%20architectures%20for%20AI%20%282009%29.pdf', 0.052999894000318216), ('http://www.scholarpedia.org/article/Deep\_Learning', 0.052999894000318216), ('http://adsabs.harvard.edu/abs/2009SchpJ...4.5947H', 0.052999894000318216), ('http://deeplearning.cs.cmu.edu/pdfs/Cybenko.pdf', 0.052999894000318216), ('http://papers.nips.cc/paper/7203-the-expressive-power-of-neural-networks-a-view-from-the-width', 0.052999894000318216), ('http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf', 0.052999894000318216), ('http://www.math.uiuc.edu/documenta/vol-ismp/52\_griewank-andreas-b.pdf', 0.052999894000318216), ('http://www.cse.msu.edu/~weng/research/CresceptronIJCNN1992.pdf', 0.052999894000318216), ('http://www.cse.msu.edu/~weng/research/CresceptronICCV1993.pdf', 0.052999894000318216), ('http://www.cse.msu.edu/~weng/research/CresceptronIJCV.pdf', 0.052999894000318216), ('http://adsabs.harvard.edu/abs/1995Sci...268.1158H', 0.052999894000318216), ('http://people.idsia.ch/~juergen/SeppHochreiter1991ThesisAdvisorSchmidhuber.pdf', 0.052999894000318216), ('http://dl.acm.org/citation.cfm?id=1895720', 0.052999894000318216), ('http://adsabs.harvard.edu/abs/2009ISPM...26...75B', 0.052999894000318216), ('http://googleresearch.blogspot.ch/2015/09/google-voice-search-faster-and-more.html', 0.052999894000318216)]

The top hub page [('https://en.wikipedia.org/wiki/Deep\_learning', 0.052999894000318154), ('http://adsabs.harvard.edu/abs/2015Natur.521..436L', 0.052999894000318154), ('http://ieeexplore.ieee.org/document/6248110/', 0.052999894000318154), ('http://adsabs.harvard.edu/abs/1996Natur.381..607O', 0.052999894000318154), ('http://research.microsoft.com/pubs/209355/DeepLearning-NowPublishing-Vol7-SIG-039.pdf', 0.052999894000318154), ('http://sanghv.com/download/soft/machine%20learning,%20artificial%20intelligence,%20mathematics%20ebooks/ML/learning%20deep%20architectures%20for%20AI%20%282009%29.pdf', 0.052999894000318154), ('http://www.scholarpedia.org/article/Deep\_Learning', 0.052999894000318154), ('http://adsabs.harvard.edu/abs/2009SchpJ...4.5947H', 0.052999894000318154), ('http://deeplearning.cs.cmu.edu/pdfs/Cybenko.pdf', 0.052999894000318154), ('http://papers.nips.cc/paper/7203-the-expressive-power-of-neural-networks-a-view-from-the-width', 0.052999894000318154), ('http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf', 0.052999894000318154), ('http://www.math.uiuc.edu/documenta/vol-ismp/52\_griewank-andreas-b.pdf', 0.052999894000318154), ('http://www.cse.msu.edu/~weng/research/CresceptronIJCNN1992.pdf', 0.052999894000318154), ('http://www.cse.msu.edu/~weng/research/CresceptronICCV1993.pdf', 0.052999894000318154), ('http://www.cse.msu.edu/~weng/research/CresceptronIJCV.pdf', 0.052999894000318154), ('http://adsabs.harvard.edu/abs/1995Sci...268.1158H', 0.052999894000318154), ('http://people.idsia.ch/~juergen/SeppHochreiter1991ThesisAdvisorSchmidhuber.pdf', 0.052999894000318154), ('http://dl.acm.org/citation.cfm?id=1895720', 0.052999894000318154), ('http://adsabs.harvard.edu/abs/2009ISPM...26...75B', 0.052999894000318154), ('http://googleresearch.blogspot.ch/2015/09/google-voice-search-faster-and-more.html', 0.052999894000318154)]

For Machine Learning:

K = 5

he top authority page [('https://www.sas.com/en\_us/insights/analytics/machine-learning.html', 0.05892556509887893), ('http://support.sas.com/training/us/paths/index.html', 0.05892556509887893), ('http://support.sas.com/rnd/index.html', 0.05892556509887893), ('http://support.sas.com/en/support-home.html', 0.05892556509887893), ('http://support.sas.com/sassamples/graphgallery/index.html', 0.05892556509887893)]

The top hub page [('https://www.sas.com/en\_us/insights/analytics/machine-learning.html', 0.05892556509887889), ('http://support.sas.com/training/us/paths/index.html', 0.05892556509887889), ('http://support.sas.com/rnd/index.html', 0.05892556509887889), ('http://support.sas.com/en/support-home.html', 0.05892556509887889), ('http://support.sas.com/sassamples/graphgallery/index.html', 0.05892556509887889)]

K = 10

The top authority page [('https://www.sas.com/en\_us/insights/analytics/machine-learning.html', 0.062136976600119964), ('http://support.sas.com/training/us/paths/index.html', 0.062136976600119964), ('http://support.sas.com/rnd/index.html', 0.062136976600119964), ('http://support.sas.com/en/support-home.html', 0.062136976600119964), ('http://support.sas.com/sassamples/graphgallery/index.html', 0.062136976600119964), ('http://support.sas.com/resources/pages/ds.html', 0.062136976600119964), ('http://support.sas.com/software/index.html', 0.062136976600119964), ('http://support.sas.com/software/products/sasstudio/index.html', 0.062136976600119964), ('http://support.sas.com/software/products/guide/index.html', 0.062136976600119964), ('http://support.sas.com/software/products/base/index.html', 0.062136976600119964)]

The top hub page [('https://www.sas.com/en\_us/insights/analytics/machine-learning.html', 0.062136976600119964), ('http://support.sas.com/training/us/paths/index.html', 0.062136976600119964), ('http://support.sas.com/rnd/index.html', 0.062136976600119964), ('http://support.sas.com/en/support-home.html', 0.062136976600119964), ('http://support.sas.com/sassamples/graphgallery/index.html', 0.062136976600119964), ('http://support.sas.com/resources/pages/ds.html', 0.062136976600119964), ('http://support.sas.com/software/index.html', 0.062136976600119964), ('http://support.sas.com/software/products/sasstudio/index.html', 0.062136976600119964), ('http://support.sas.com/software/products/guide/index.html', 0.062136976600119964), ('http://support.sas.com/software/products/base/index.html', 0.062136976600119964)]

K = 20

The top authority page [('https://www.sas.com/en\_us/insights/analytics/machine-learning.html', 0.05070201265633947), ('http://support.sas.com/training/us/paths/index.html', 0.05070201265633947), ('http://support.sas.com/rnd/index.html', 0.05070201265633947), ('http://support.sas.com/en/support-home.html', 0.05070201265633947), ('http://support.sas.com/sassamples/graphgallery/index.html', 0.05070201265633947), ('http://support.sas.com/resources/pages/ds.html', 0.05070201265633947), ('http://support.sas.com/software/index.html', 0.05070201265633947), ('http://support.sas.com/software/products/sasstudio/index.html', 0.05070201265633947), ('http://support.sas.com/software/products/guide/index.html', 0.05070201265633947), ('http://support.sas.com/software/products/base/index.html', 0.05070201265633947), ('http://support.sas.com/software/products/va/index.html', 0.05070201265633947), ('http://support.sas.com/software/products/stat/index.html', 0.05070201265633947), ('http://support.sas.com/software/products/miner/index.html', 0.05070201265633947), ('http://support.sas.com/downloads/index.htm', 0.05070201265633947), ('http://support.sas.com/en/sas-administrators.html', 0.05070201265633947), ('http://support.sas.com/ctx/supportform/createForm', 0.05070201265633947), ('http://support.sas.com/en/product-resources.html', 0.05070201265633947), ('http://support.sas.com/documentation/installcenter/index.html', 0.05070201265633947), ('http://support.sas.com/resources/sysreq/index.html', 0.05070201265633947), ('http://support.sas.com/techsup/license/index.html', 0.05070201265633947)]

The top hub page [('https://www.sas.com/en\_us/insights/analytics/machine-learning.html', 0.05070201265633947), ('http://support.sas.com/training/us/paths/index.html', 0.05070201265633947), ('http://support.sas.com/rnd/index.html', 0.05070201265633947), ('http://support.sas.com/en/support-home.html', 0.05070201265633947), ('http://support.sas.com/sassamples/graphgallery/index.html', 0.05070201265633947), ('http://support.sas.com/resources/pages/ds.html', 0.05070201265633947), ('http://support.sas.com/software/index.html', 0.05070201265633947), ('http://support.sas.com/software/products/sasstudio/index.html', 0.05070201265633947), ('http://support.sas.com/software/products/guide/index.html', 0.05070201265633947), ('http://support.sas.com/software/products/base/index.html', 0.05070201265633947), ('http://support.sas.com/software/products/va/index.html', 0.05070201265633947), ('http://support.sas.com/software/products/stat/index.html', 0.05070201265633947), ('http://support.sas.com/software/products/miner/index.html', 0.05070201265633947), ('http://support.sas.com/downloads/index.htm', 0.05070201265633947), ('http://support.sas.com/en/sas-administrators.html', 0.05070201265633947), ('http://support.sas.com/ctx/supportform/createForm', 0.05070201265633947), ('http://support.sas.com/en/product-resources.html', 0.05070201265633947), ('http://support.sas.com/documentation/installcenter/index.html', 0.05070201265633947), ('http://support.sas.com/resources/sysreq/index.html', 0.05070201265633947), ('http://support.sas.com/techsup/license/index.html', 0.0507020126563

Source Code:

from urllib.request import urlopen

from bs4 import BeautifulSoup

import re

import requests

from graphviz import Digraph

import networkx as nx

import math

import heapq

avoid\_domain = ['facebook', 'twitter', 'linkedin', 'youtube', 'doubleclick', 'ads', '.png', '.jpg', '.svg']

rootSet = []

neighbor = []

G = nx.Graph()

DG = nx.DiGraph()

web = []

def build\_Inneighbor\_set(root\_set, k):

count = 0

#url = link

url = "https://search.yahoo.com/search?q=link:"+root\_set

html = requests.get(url)

soup = BeautifulSoup(html.text, 'lxml')

links = soup.find\_all('a', attrs={'href': re.compile("^http://")} ,href = True)

for link in links:

if link.get('class') and "ac-algo" in link['class']:

neighbor.append([link['href'], root\_set])

G.add\_edges\_from([(neighbor[count][0], root\_set)])

count = count + 1

if count== k/2:

break

#print (incidents)

return

def build\_Outneighbor\_set(root\_set, k):

count = 0

html = requests.get(root\_set)

soup = BeautifulSoup(html.text, 'lxml')

links = soup.find\_all('a', attrs={'href': re.compile("^http://")} )

for link in links:

neighbor.append([root\_set, link['href']])

G.add\_edges\_from([(root\_set, neighbor[count][1])])

count = count + 1

if count == k:

break

#out\_neighbor = random.choice(neighbor, k)

#print(neighboors)

return

def build\_root\_set(search\_term):

count = 1

header = {'User-Agent': 'Mozilla/5.0 (X11; U; Linux i686) Gecko/20071127 Firefox/2.0.0.11 Chrome/53.0.2785.143 Safari/537.36'}

url = "https://www.google.com/search?q="+search\_term+"&num=60"

raw\_page = requests.get(url, headers = header).text

results = re.findall(r'(?<=<h3 class="r"><a href="/url\?q=).\*?(?=&amp)', str(raw\_page))

for line in results:

domain\_name = line.split('.')[1].split('.')[0]

if domain\_name not in avoid\_domain and count <= 30:

rootSet.append(line)

count = count + 1

#print(count)

#print (len(rootSet))

return rootSet

class HITSIterator:

def \_\_init\_\_(self, G):

self.errorrate = 0.00001

self.max\_iterations = 100

self.graph = DG

self.hub = {}

self.authority = {}

for node in self.graph.nodes():

self.hub[node] = 1

self.authority[node] = 1

#print (self.authority)

def hits(self, k):

flag = False

for i in range(self.max\_iterations):

change = 0.0

norm = 0

tmp = {}

tmp = self.authority.copy()

for node in self.graph.nodes():

self.authority[node] = 0

for incident\_page in self.graph.in\_edges(nbunch = None, data = False):

#print([incident\_page[1]])

#print(self.authority[node])

self.authority[node] += self.hub[incident\_page[0]]

norm += pow(self.authority[node], 2)

norm = math.sqrt(norm)

for node in self.graph.nodes():

self.authority[node] /= norm

change += abs(tmp[node] - self.authority[node])

#=======================================================

#===========hub=======================================

norm = 0

tmp = self.hub.copy()

for node in self.graph.nodes():

self.hub[node] = 0

for neighbor\_page in self.graph.out\_edges(nbunch = None, data = False):

self.hub[node] += self.authority[neighbor\_page[1]]

norm += pow(self.hub[node], 2)

norm = math.sqrt(norm)

for node in self.graph.nodes():

self.hub[node] /= norm

change += abs(tmp[node] - self.hub[node])

#print("authority\n", self.authority)

#print("hub\n", self.hub)

if change < self.errorrate:

flag = True

break

if flag:

print("finished in %s iterations" % (i + 1))

#for i in range (0, k):

print("The top authority page", heapq.nlargest(k,self.authority.items(), key= lambda x:x[1]))

print("The top hub page", heapq.nlargest(k,self.hub.items(), key= lambda x:x[1]))

def main():

root\_set = build\_root\_set('machine+learning')

k =20 #input()

for i in range (0,30):

web = root\_set[i]

#print("root\_web",web)

#print("=========================================\n")

build\_Outneighbor\_set(web, k)

build\_Inneighbor\_set(web, k)

list(nx.connected\_components(G))

#print(list(nx.connected\_components(G)))

#print(neighbor)

#print("=========================================\n")

for i in neighbor:

DG.add\_edges\_from([(i[0],i[1])])

#node, neighbors, data= DG.in\_edges(nbunch = None, data = False)

#print(DG.out\_edges(nbunch = None, data = False))

#print(DG.nodes())

hits = HITSIterator(G)

hits.hits(k)

if \_\_name\_\_ == '\_\_main\_\_':

main()