Assignment 8 Report

Professor Nelson

By Tyler Medina

4/13/17

Part 1

1.1 Getting 100 Blogs

In getBlogs.py, I used BeautifulSoup to parse the blogs and get the uris.

```
http://f-measure.blogspot.com/feeds/posts/default?max-results=200
http://ws-dl.blogspot.com/feeds/posts/default?max-results=200
http://homeschoolingmomjournel.blogspot.com/feeds/posts/default?max-results=1000
http://lifelonglearningadventures.blogspot.com/feeds/posts/default?max-results=1000
http://asoleperspective.blogspot.com/feeds/posts/default?max-results=1000
http://aquiverfull-christy.blogspot.com/feeds/posts/default?max-results=1000
http://adamedwardsinparis.blogspot.com/feeds/posts/default?max-results=1000
http://barcelonadreamin.blogspot.com/feeds/posts/default?max-results=1000
http://firstsundaything.blogspot.com/feeds/posts/default?max-results=1000
http://pastorronking.blogspot.com/feeds/posts/default?max-results=1000
http://elderstringfellow.blogspot.com/feeds/posts/default?max-results=1000
https://pureanthology.blogspot.com/feeds/posts/default?max-results=1000
http://feeds.feedburner.com/blogspot/ehhaw?max-results=1000
http://archibaldinsc.blogspot.com/feeds/posts/default?max-results=1000
http://ielingslife.blogspot.com/feeds/posts/default?max-results=1000
http://adarobinson.blogspot.com/feeds/posts/default?max-results=1000
http://drolleryandpathos.blogspot.com/feeds/posts/default?max-results=1000
http://mylittlethoughtsandhisbigones.blogspot.com/feeds/posts/default?max-results=1000
http://revmiriam.blogspot.com/feeds/posts/default?max-results=1000
http://reverendref.blogspot.com/feeds/posts/default?max-results=1000
http://melyndamac.blogspot.com/feeds/posts/default?max-results=1000
http://therosenortcommunitychurch.blogspot.com/feeds/posts/default?max-results=1000
http://bethslineuponline.blogspot.com/feeds/posts/default?max-results=1000
http://godayame.blogspot.com/feeds/posts/default?max-results=1000
http://www.karmannscotulla.com/feeds/posts/default?max-results=1000
http://feeds.feedburner.com/ASeriesOfLinks?max-results=1000
http://mondaymorningbibleblog.blogspot.com/feeds/posts/default?max-results=1000
http://mauryn143.blogspot.com/feeds/posts/default?max-results=1000
http://thegirlinrippedjeans.blogspot.com/feeds/posts/default?max-results=1000
http://adventuresonbirchcreekroad.blogspot.com/feeds/posts/default?max-results=1000
http://tpaullee.blogspot.com/feeds/posts/default?max-results=1000
http://valiantrace.blogspot.com/feeds/posts/default?max-results=1000
http://hopeforthehormonal.blogspot.com/feeds/posts/default?max-results=1000
```

http://feeds.feedburner.com/shorttermmissiontrips?max-results=1000
http://fonnymichael.blogspot.com/feeds/posts/default?max-results=1000
http://abucketofgrace.blogspot.com/feeds/posts/default?max-results=1000
http://awarian-raphael.blogspot.com/feeds/posts/default?max-results=1000
http://awjourney7.blogspot.com/feeds/posts/default?max-results=1000
http://carolweeks.blogspot.com/feeds/posts/default?max-results=1000
http://tru-treasures.blogspot.com/feeds/posts/default?max-results=1000
http://just-for-fun-jess.blogspot.com/feeds/posts/default?max-results=1000
http://cheneyscene.blogspot.com/feeds/posts/default?max-results=1000
http://emelyne-mei.blogspot.com/feeds/posts/default?max-results=1000
http://rhetoricandcivilization.blogspot.com/feeds/posts/default?max-results=1000
http://randomrevhd.blogspot.com/feeds/posts/default?max-results=1000
http://findinggracewithin.blogspot.com/feeds/posts/default?max-results=1000
http://feeds.feedburner.com/HouseOfGlory?max-results=1000
http://altion-days.blogspot.com/feeds/posts/default?max-results=1000
http://bittylittlekisses.blogspot.com/feeds/posts/default?max-results=1000
http://storiesofsimcha.blogspot.com/feeds/posts/default?max-results=1000
http://catherinefangs.blogspot.com/feeds/posts/default?max-results=1000
http://livinglifetoinspireothers.blogspot.com/feeds/posts/default?max-results=1000
http://familyfaithincyberspace.blogspot.com/feeds/posts/default?max-results=1000
http://wouldyouliketosingasong.blogspot.com/feeds/posts/default?max-results=1000
http://politicallymoderatejesusradical.blogspot.com/feeds/posts/default?max-results=1000
http://wisewomanalchemy.blogspot.com/feeds/posts/default?max-results=1000
http://courtneydavine.blogspot.com/feeds/posts/default?max-results=1000
https://joyful-peace.blogspot.com/feeds/posts/default?max-results=1000
http://ceeharmon.blogspot.com/feeds/posts/default?max-results=1000
http://www.constance.nu/feeds/posts/default?max-results=1000
http://seeinhislight.blogspot.com/feeds/posts/default?max-results=1000
http://lou52.blogspot.com/feeds/posts/default?max-results=1000
http://kerensagrayjazz.blogspot.com/feeds/posts/default?max-results=1000
http://reformingpastor.blogspot.com/feeds/posts/default?max-results=1000
http://jez-carina-simpson.blogspot.com/feeds/posts/default?max-results=1000
http://ramblingmystic.blogspot.com/feeds/posts/default?max-results=1000

```
http://understandlove11.blogspot.com/feeds/posts/default?max-results=1000
http://ycahpadayao.blogspot.com/feeds/posts/default?max-results=1000
http://realitiesandrealizations.blogspot.com/feeds/posts/default?max-results=1000
http://feeds.feedburner.com/blogspot/pRBKnI?max-results=1000
http://rachelssimplelife.blogspot.com/feeds/posts/default?max-results=1000
http://thefiresofsekhmet.blogspot.com/feeds/posts/default?max-results=1000
http://wabisabiwords.blogspot.com/feeds/posts/default?max-results=1000
http://footprintsonthecamino.blogspot.com/feeds/posts/default?max-results=1000
http://clmroma.blogspot.com/feeds/posts/default?max-results=1000
http://hopeabel.blogspot.com/feeds/posts/default?max-results=1000
http://synergiesforwiseliving.blogspot.com/feeds/posts/default?max-results=1000
http://theupsidedownpastorswife.blogspot.com/feeds/posts/default?max-results=1000
http://poetry4adalia.blogspot.com/feeds/posts/default?max-results=1000
http://tanzaniasurgery.blogspot.com/feeds/posts/default?max-results=1000
http://onionquilter.blogspot.com/feeds/posts/default?max-results=1000
http://feeds.feedburner.com/blogspot/xilxz?max-results=1000
http://truewomenblog.blogspot.com/feeds/posts/default?max-results=1000
http://nataliesfhespot.blogspot.com/feeds/posts/default?max-results=1000
http://ourlittleblueheaven.blogspot.com/feeds/posts/default?max-results=1000
http://remnantsoflight.blogspot.com/feeds/posts/default?max-results=1000
http://builtbystory.blogspot.com/feeds/posts/default?max-results=1000
http://theblueeyedbunch.blogspot.com/feeds/posts/default?max-results=1000
http://elderohuishibashi-fiji.blogspot.com/feeds/posts/default?max-results=1000
http://vanberean.blogspot.com/feeds/posts/default?max-results=1000
http://eknowledgehub.blogspot.com/feeds/posts/default?max-results=1000
http://jprice1010.blogspot.com/feeds/posts/default?max-results=1000
http://dundetails.blogspot.com/feeds/posts/default?max-results=1000
http://thejoyofhope21.blogspot.com/feeds/posts/default?max-results=1000
http://timandlizdecker.blogspot.com/feeds/posts/default?max-results=1000
http://stoandthecity.blogspot.com/feeds/posts/default?max-results=1000
http://ldsliberty.blogspot.com/feeds/posts/default?max-results=1000
https://gistcitiafrica.blogspot.com/feeds/posts/default?max-results=1000
http://aroundthehouseministries.blogspot.com/feeds/posts/default?max-results=1000
```

1.2 Making a Blog-Term Matrix

To be able to make a blog-term matrix, I used the getwordcounts() and getwords() functions in the collective intelligence textbook. Getwordcounts() gets the title and word counts for the RSS feed. The getwords() function strips the html and returns the words as a list.

```python
]def getwordcounts(url):
    # Parse the feed
    d=feedparser.parse(url)
    wc={}

    # Loop over all the entries
    for e in d.entries:
        if 'summary' in e: summary=e.summary
        else: summary=e.description

        # Extract a list of words
        words=getwords(e.title+' '+summary)
        for word in words:
            wc.setdefault(word,0)
            wc[word]+=1
    return d.feed.title,wc

]def getwords(html):
    # Remove all the HTML tags
    txt=re.compile(r'<[^>]+>').sub('',html)

    # Split words by all non-alpha characters
    words=re.compile(r'[^A-Z^a-z]+').split(txt)

    # Convert to lowercase
    return [word.lower() for word in words if word!='']
```

To get a list of words we limit the amount of times the word occurs between 10 percent and fifty percent. This ensures that we throw out very specific words that'll have title value and stop words such as "the". With these bounds, we can craft a blog-term matrix.

```python
wordlist=[]
for w,bc in apcount.items():
    frac=float(bc)/len(feedlist)
    if frac>0.1 and frac<0.5:
        wordlist.append(w)
        if len(wordlist) >= 1000:
            break
```

Blog	hanging	bringing		wednesday		woods		cooking	china		climbed	golden		circumstances			locked		absolute			welcomed			fix	effects	encourage		husbands		
ycah	0	0	3	0	2	0	0	0	1	0	0	1	0	0	9	0	2	3	0	1	0	0	2	3	1	5	0	0	0	1	0	8	0	2	0	0	0	0	0
OM Arts International	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Synergies for Wise Living	0	2	0	2	2	0	0	0	0	0	0	0	0	1	0	0	0	0	1	3	3	0	0	0	0	1	0	0	1	0	1	0	0	0					
Unpluckable	5	7	2	0	7	0	0	1	1	0	2	0	2	1	6	3	1	1	2	1	0	0	0	0	0	0	0	0	0	4	0	0	0	5	0	0	0	4	
Gistcity Africa Media	0	1	2	0	1	4	0	3	1	0	1	0	0	0	1	1	0	1	1	1	1	4	5	0	1	8	2	0	0	8	0	0	0	0	2				
Rev. H-D's random ramblings, really!		2	15	7	1	1	1	0	2	7	8	2	9	1	0	3	1	0	23	108	4	0	1	0	1	3	1	0	1	2	1	0							
Reforming Pastor	1	2	0	1	0	0	0	0	7	0	2	0	7	3	5	0	6	5	3	7	2	0	2	0	2	0	0	0	0	1	0	0	0	0	1	1			
The Blue Eyed Bunch	1	2	1	0	0	0	4	2	1	2	0	1	6	0	2	0	0	0	0	0	0	0	0	0	8	1	0	0	0	0	1	0	5	1	1				
A Van Berean Blog	1	1	0	0	0	0	0	1	3	0	1	1	1	0	7	7	0	0	27	2	2	0	0	0	6	0	0	2	7	5	1	0	0	0	0				
Web Science and Digital Libraries Research Group		0	12	15	0	4	3	0	0	1	3	3	8	2	8	22	0	0	20	0	18	2	5	1	0	2	2	0	6										
Adalia's Poetry	1	1	0	1	0	0	0	2	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0				
Francis X Archibald	7	1	6	1	0	6	0	0	0	5	0	2	0	0	1	0	0	4	0	2	3	10	0	0	0	2	2	1	0	4	0	0	1	4	10	0			
Finding Grace Within	1	0	3	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1				
The Upside-Down Pastor's Wife		8	0	1	0	0	0	2	1	6	2	1	0	0	1	1	0	0	4	1	4	0	0	0	1	0	0	4	0	1	0	0	1	0					
New Hope Family Worship Center	0	2	1	0	0	0	2	0	0	0	0	0	1	0	5	2	0	19	2	3	0	0	0	0	4	0	0	0	1	4	1	0	1						
in someone else's words	3	1	0	0	0	0	0	3	5	0	1	0	2	1	0	0	1	2	2	2	0	0	4	0	2	0	2	1	0	0	0	0	2	0					
Mystical Ramblings	0	1	0	0	0	0	0	1	4	0	2	0	0	1	0	0	0	0	3	1	0	0	0	0	0	2	2	0	0	2	0	1	1	0	0	0			
Goddess Alchemy~A Priestess Journey	0	10	0	0	0	2	0	6	2	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	31	0	0								
.... Altion Days	0	1	4	1	0	2	0	1	1	0	4	3	1	2	0	1	0	3	14	7	2	1	0	0	0	0	0	0	1	2	0	0	0	1	0				
Wabi Sabi Words	3	1	2	0	2	0	0	2	4	2	2	0	1	1	2	0	0	1	0	2	1	1	0	1	1	0	0	0	0	3	0	5	1	3	1	0	0		
tpaullee		3	0	0	2	1	117	0	0	4	1	0	0	4	0	3	0	0	3	0	2	0	1	0	0	0	4	4	0	3	2	2	1	3	8	4	1	0	0
Our little blue heaven	0	1	4	3	2	0	2	0	6	1	1	1	2	0	0	1	0	2	0	3	2	0	0	3	1	2	0	0	3	0	1	0	0	1	0				
ekh - e-Knowledge Hub	0	2	1	0	2	3	2	0	3	1	11	0	1	4	1	0	0	5	0	6	0	5	0	1	0	2	0	0	2	5	1	7	1	2	1				
The Family Home Evening Spot		2	0	0	0	1	0	0	1	0	0	0	0	1	0	5	0	0	1	4	1	0	0	0	1	0	3	11	2	0	1	0	0	2					
leaning into life	7	1	2	5	1	1	1	0	3	0	0	0	4	0	2	0	0	3	0	3	0	0	0	4	0	0	10	0	0	0	1	7	1	4	0	2			
Valiant	1	3	13	0	0	0	0	4	20	0	9	0	0	0	0	12	0	64	11	11	4	0	7	0	0	11	0	0	3	6	0	0	3	0	0	0	0	0	0

Part 2

2.1 Creating a Dendrogram

The Collective Intelligence textbook includes the functions to create the dendrogram. By using the hcluster() function, this ensures that the clusters recursively compare themselves with each other before merging. This process will continue until only one cluster is rest.

```python
]def hcluster(rows,distance=pearson):
   distances={}
   currentclustid=-1

   # Clusters are initially just the rows
   clust=[bicluster(rows[i],id=i) for i in range(len(rows))]

   while len(clust)>1:
     lowestpair=(0,1)
     closest=distance(clust[0].vec,clust[1].vec)

     # loop through every pair looking for the smallest distance
     for i in range(len(clust)):
       for j in range(i+1,len(clust)):
         # distances is the cache of distance calculations
         if (clust[i].id,clust[j].id) not in distances:
           distances[(clust[i].id,clust[j].id)]=distance(clust[i].vec,clust[j].vec)

         d=distances[(clust[i].id,clust[j].id)]

         if d<closest:
           closest=d
           lowestpair=(i,j)

     # calculate the average of the two clusters
     mergevec=[
     (clust[lowestpair[0]].vec[i]+clust[lowestpair[1]].vec[i])/2.0
     for i in range(len(clust[0].vec))]

     # create the new cluster
     newcluster=bicluster(mergevec,left=clust[lowestpair[0]],
                          right=clust[lowestpair[1]],
                          distance=closest,id=currentclustid)
```
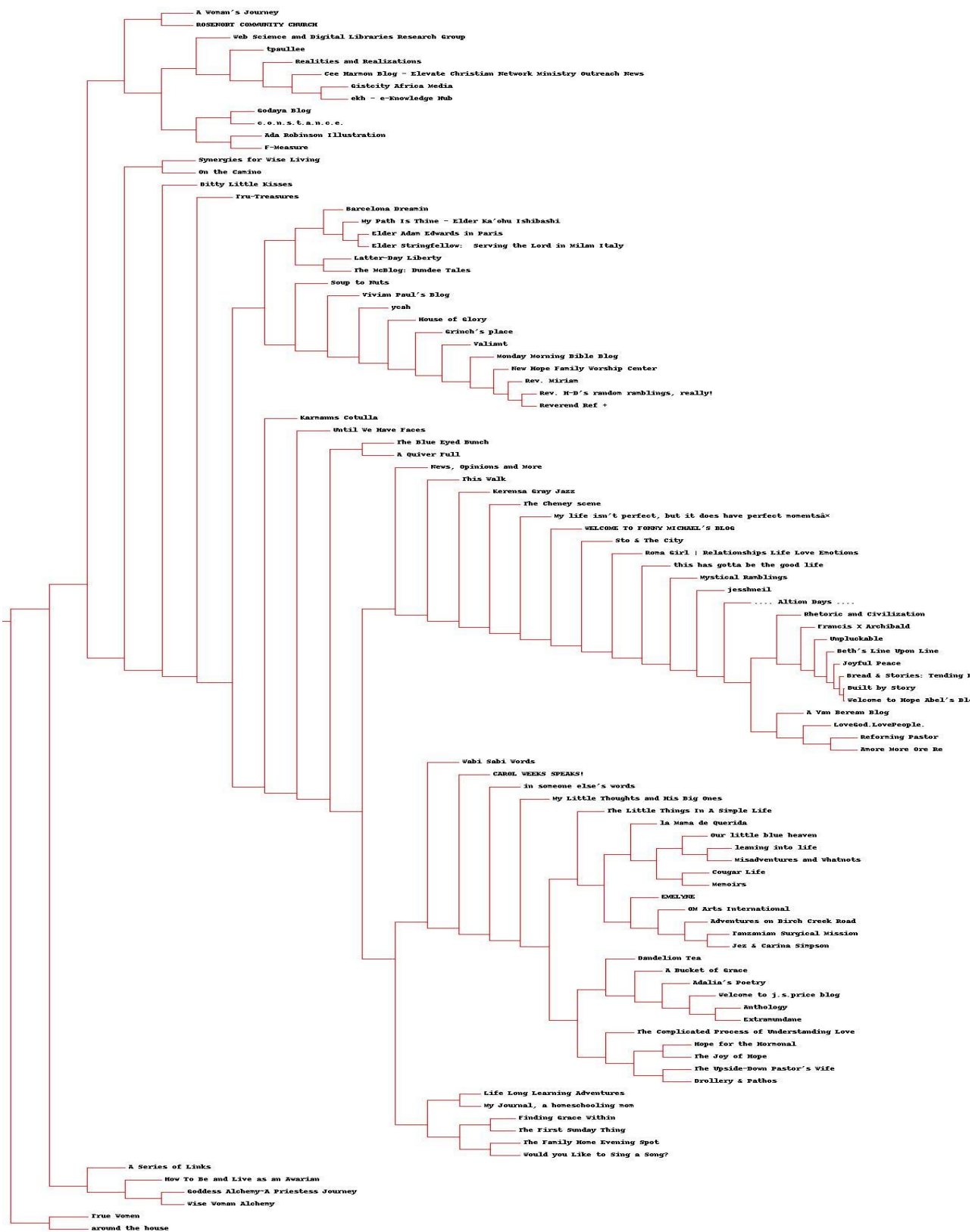
A Woman's Journey
ROSENORT COMMUNITY CHURCH
Web Science and Digital Libraries Research Group
tpaullee
Realities and Realizations
Cee Harmon Blog - Elevate Christian Network Ministry Outreach News
Gistcity Africa Media
ekh - e-Knowledge Hub
Godaya Blog
c.o.n.s.t.a.n.c.e.
Ada Robinson Illustration
F-Measure
Synergies for Wise Living
On the Camino
Bitty Little Kisses
Tru-Treasures
Barcelona Dreamin
My Path Is Thine - Elder Ka'ohu Ishibashi
Elder Adam Edwards in Paris
Elder Stringfellow: Serving the Lord in Milan Italy
Latter-Day Liberty
The McBlog: Dundee Tales
Soup to Nuts
Vivian Paul's Blog
ycah
House of Glory
Grinch's place
Valiant
Monday Morning Bible Blog
New Hope Family Worship Center
Rev. Miriam
Rev. M-D's random ramblings, really!
Reverend Ref +
Karmamns Cotulla
Until We Have Faces
The Blue Eyed Bunch
A Quiver Full
News, Opinions and More
This Walk
Kerensa Gray Jazz
The Cheney scene
My life isn't perfect, but it does have perfect momentsá×
WELCOME TO FONNY MICHAEL'S BLOG
Sto & The City
Roma Girl | Relationships Life Love Emotions
this has gotta be the good life
Mystical Ramblings
jesshmeil
.... Altion Days ....
Rhetoric and Civilization
Francis X Archibald
Unpluckable
Beth's Line Upon Line
Joyful Peace
Bread & Stories: Tending
Built by Story
Welcome to Hope Abel's Bl
A Van Berean Blog
LoveGod.LovePeople.
Reforming Pastor
Amore More Ore Re
Wabi Sabi Words
CAROL WEEKS SPEAKS!
in someone else's words
My Little Thoughts and His Big Ones
The Little Things In A Simple Life
la Mama de Querida
Our little blue heaven
leaning into life
Misadventures and Whatnots
Cougar Life
Memoirs
EMELYNE
OM Arts International
Adventures on Birch Creek Road
Tanzanian Surgical Mission
Jez & Carina Simpson
Dandelion Tea
A Bucket of Grace
Adalia's Poetry
Welcome to j.s.price blog
Anthology
Extramundane
The Complicated Process of Understanding Love
Hope for the Hormonal
The Joy of Hope
The Upside-Down Pastor's Wife
Drollery & Pathos
Life Long Learning Adventures
My Journal, a homeschooling mom
Finding Grace Within
The First Sunday Thing
The Family Home Evening Spot
Would you Like to Sing a Song?
A Series of Links
How To Be and Live as an Awarian
Goddess Alchemy~A Priestess Journey
Wise Woman Alchemy
True Women
around the house

Part 3

3.1 Kcluster

Using k-means instead of hierarchical is similar into part 2, but we use the kcluster()
function instead of the hcluster() function.

```python
def kcluster(rows,distance=pearson,k=4):
  # Determine the minimum and maximum values for each point
  ranges=[(min([row[i] for row in rows]),max([row[i] for row in rows]))
  for i in range(len(rows[0]))]
  # Create k randomly placed centroids
  clusters=[[random.random()*(ranges[i][1]-ranges[i][0])+ranges[i][0]
  for i in range(len(rows[0]))] for j in range(k)]
  lastmatches=None
  for t in range(100):
    print 'Iteration %d' % t
    bestmatches=[[] for i in range(k)]
    # Find which centroid is the closest for each row
    for j in range(len(rows)):
      row=rows[j]
      bestmatch=0
      for i in range(k):
        d=distance(clusters[i],row)
        if d<distance(clusters[bestmatch],row): bestmatch=i
      bestmatches[bestmatch].append(j)
    # If the results are the same as last time, this is complete
    if bestmatches==lastmatches: break
    lastmatches=bestmatches
    # Move the centroids to the average of their members
    for i in range(k):
      avgs=[0.0]*len(rows[0])
      if len(bestmatches[i])>0:
        for rowid in bestmatches[i]:
          for m in range(len(rows[rowid])):
            avgs[m]+=rows[rowid][m]
        for j in range(len(avgs)):
          avgs[j]/=len(bestmatches[i])
        clusters[i]=avgs
  return bestmatches
```

length : 8561   lines :

3.2 Getting the centroids and iterations

I pass k=5 as a parameter to get the number of iterations to get 5 clusters.

```python
(blognames, words, data) = clusters.readfile('blogdata.txt')




kclust=clusters.kcluster(data,k=5)
#kclust=clusters.kcluster(data,k=10)
#kclust=clusters.kcluster(data,k=20)
print "\n\n"+str([blognames[r] for r in kclust[0]])
print "\n\n"+str([blognames[r] for r in kclust[1]])
print "\n\n"+str([blognames[r] for r in kclust[2]])
print "\n\n"+str([blognames[r] for r in kclust[3]])
print "\n\n"+str([blognames[r] for r in kclust[4]])
"""
print "\n\n"+str([blognames[r] for r in kclust[5]])
print "\n\n"+str([blognames[r] for r in kclust[6]])
print "\n\n"+str([blognames[r] for r in kclust[7]])
print "\n\n"+str([blognames[r] for r in kclust[8]])
print "\n\n"+str([blognames[r] for r in kclust[9]])
"""
"""
print "\n\n"+str([blognames[r] for r in kclust[10]])
print "\n\n"+str([blognames[r] for r in kclust[11]])
print "\n\n"+str([blognames[r] for r in kclust[12]])
print "\n\n"+str([blognames[r] for r in kclust[13]])
print "\n\n"+str([blognames[r] for r in kclust[14]])
print "\n\n"+str([blognames[r] for r in kclust[15]])
print "\n\n"+str([blognames[r] for r in kclust[16]])
print "\n\n"+str([blognames[r] for r in kclust[17]])
print "\n\n"+str([blognames[r] for r in kclust[18]])
print "\n\n"+str([blognames[r] for r in kclust[19]])
```

When k = 5, it takes 4 iterations.

```
Iteration 0
Iteration 1
Iteration 2
Iteration 3

["Adalia's Poetry", "The Upside-Down Pastor's Wife", 'Goddess Alchemy~A Priestess Journey', 'Our little blue heaven', "A Woman's Journey", 'Welcome to j.s.price blog',
'How To Be and Live as an Awarian', 'Misadventures and Whatnots', 'Hope for the Hormonal', 'Memoirs', 'Godaya Blog', 'Dandelion Tea', 'Anthology', 'Wise Woman Alchemy',
'Extramundane', 'Drollery & Pathos', 'A Bucket of Grace']

['ycah', 'Synergies for Wise Living', "Rev. H-D's random ramblings, really!", 'Reforming Pastor', 'A Van Berean Blog', 'New Hope Family Worship Center', '.... Altion Da
ys ....', 'ekh - e-Knowledge Hub', 'The Family Home Evening Spot', 'Valiant', 'Amore More Ore Re', 'Life Long Learning Adventures', 'True Women', 'A Quiver Full', 'this
has gotta be the good life', 'Bitty Little Kisses', 'Until We Have Faces', 'The First Sunday Thing', "Grinch's place", 'Reverend Ref +', 'Monday Morning Bible Blog',
House of Glory', 'Latter-Day Liberty', 'LoveGod.LovePeople.', 'This Walk', 'Roma Girl | Relationships Life Love Emotions', 'CAROL WEEKS SPEAKS!', 'Rhetoric and Civiliza
tion', 'Vivian Paul's Blog', 'c.o.n.s.t.a.n.c.e.', 'WELCOME TO FONNY MICHAEL'S BLOG', 'Rev. Miriam', 'A Series of Links', 'The Complicated Process of Understanding Love
', 'My Journal, a homeschooling mom']

['Unpluckable', 'Gistcity Africa Media', 'Francis X Archibald', 'Finding Grace Within', 'Mystical Ramblings', 'Wabi Sabi Words', 'leaning into life', 'Bread & Stories:
Tending Family Faith in Cyberspace', 'Would you Like to Sing a Song?', 'Joyful Peace', 'Realities and Realizations', 'The Joy of Hope', 'Built by Story', 'around the ho
use', 'F-Measure', 'jesshneil', 'Welcome to Hope Abel's Blog', "Beth's Line Upon Line']

['Web Science and Digital Libraries Research Group', 'Cee Harmon Blog - Elevate Christian Network Ministry Outreach News']

['OM Arts International', 'The Blue Eyed Bunch', "in someone else's words", 'tpaullee', 'Elder Adam Edwards in Paris', 'Sto & The City', 'The Little Things In A Simple
Life', 'My Little Thoughts and His Big Ones', 'EMELYNE', 'Ada Robinson Illustration', 'ROSENORT COMMUNITY CHURCH', 'The Cheney scene', 'Adventures on Birch Creek Road',
'Soup to Nuts', 'My Path Is Thine - Elder Ka'ohu Ishibashi', 'On the Camino', 'My life isn't perfect, but it does have perfect moments\xe2\x9d\xa4', 'Elder Stringfello
w: Serving the Lord in Milan Italy', 'Barcelona Dreamin', 'la Mama de Querida', 'Cougar Life', 'Kerensa Gray Jazz', 'Karmanns Cotulla', 'News, Opinions and More', 'Tan
zanian Surgical Mission', 'Jez & Carina Simpson', 'Tru-Treasures', 'The McBlog: Dundee Tales']
```

When k=10, it takes 9 iterations.

```
Iteration 0
Iteration 1
Iteration 2
Iteration 3
Iteration 4
Iteration 5
Iteration 6
Iteration 7
Iteration 8

['OM Arts International', 'Gistcity Africa Media', 'ekh - e-Knowledge Hub', 'The First Sunday Thing', 'Cee Harmon Blog - Elevate Christian Network Ministry Outreach New
s', 'Realities and Realizations', 'around the house']

['Synergies for Wise Living', "Adalia's Poetry", 'Goddess Alchemy~A Priestess Journey', 'The Family Home Evening Spot', 'Welcome to j.s.price blog', 'On the Camino', "M
y life isn't perfect, but it does have perfect moments\xe2\x9d\xa4', 'Godaya Blog', 'Dandelion Tea', 'Wise Woman Alchemy', 'c.o.n.s.t.a.n.c.e.']

[]

['ycah', 'The Blue Eyed Bunch', 'tpaullee', 'Life Long Learning Adventures', 'Elder Adam Edwards in Paris', 'EMELYNE', 'Ada Robinson Illustration', 'The Cheney scene',
'Adventures on Birch Creek Road', 'Latter-Day Liberty', 'Soup to Nuts', 'My Path Is Thine - Elder Ka'ohu Ishibashi', 'Elder Stringfellow: Serving the Lord in Milan Ita
ly', 'Barcelona Dreamin', 'la Mama de Querida', 'Cougar Life', 'Tanzanian Surgical Mission', 'Jez & Carina Simpson', 'F-Measure', 'Tru-Treasures', 'The McBlog: Dundee T
ales']

['Unpluckable', 'A Van Berean Blog', 'Francis X Archibald', 'Finding Grace Within', 'Mystical Ramblings', '.... Altion Days ....', 'Wabi Sabi Words', 'Bread & Stories:
Tending Family Faith in Cyberspace', 'A Quiver Full', 'this has gotta be the good life', 'Would you Like to Sing a Song?', 'Joyful Peace', 'Built by Story', 'Rhetoric a
nd Civilization', 'jesshneil', 'Welcome to Hope Abel's Blog', 'My Journal, a homeschooling mom', "Beth's Line Upon Line']
```

When k=20, it takes 6 iterations.

```
Iteration 0
Iteration 1
Iteration 2
Iteration 3
Iteration 4
Iteration 5

["Adalia's Poetry", 'How To Be and Live as an Awarian', 'Kerensa Gray Jazz', 'Dandelion Tea', 'Wise Woman Alchemy']

['Web Science and Digital Libraries Research Group']

['Sto & The City']

['New Hope Family Worship Center', 'Welcome to j.s.price blog', 'Rev. Miriam']

[]

[]

['Anthology', 'Extramundane']

['ycah', 'tpaullee', 'Bitty Little Kisses', 'My Little Thoughts and His Big Ones', 'Until We Have Faces', 'Ada Robinson Illustration', 'Adventures on Birch Creek Road',
'LoveGod.LovePeople.', 'Barcelona Dreamin', 'la Mama de Querida', 'This Walk', 'Cougar Life', 'Karmanns Cotulla', 'CAROL WEEKS SPEAKS!', 'News, Opinions and More', 'F-
Measure', 'Tru-Treasures', 'A Bucket of Grace']
```

Part 4

4.1 Creating MDS

I used the scaledown() draw2d() function in the collective intelligence book to create the MDS style display. This creates a 2D representation of the data set.

```python
def scaledown(data,distance=pearson,rate=0.01):
    n=len(data)

    # The real distances between every pair of items
    realdist=[[distance(data[i],data[j]) for j in range(n)]
             for i in range(0,n)]

    # Randomly initialize the starting points of the locations in 2D
    loc=[[random.random(),random.random()] for i in range(n)]
    fakedist=[[0.0 for j in range(n)] for i in range(n)]

    lasterror=None
    for m in range(0,1000):
      # Find projected distances
      for i in range(n):
        for j in range(n):
          fakedist[i][j]=sqrt(sum([pow(loc[i][x]-loc[j][x],2)
                              for x in range(len(loc[i]))]))

      # Move points
      grad=[[0.0,0.0] for i in range(n)]

      totalerror=0
      for k in range(n):
        for j in range(n):
          if j==k: continue
          # The error is percent difference between the distances
          errorterm=(fakedist[j][k]-realdist[j][k])/realdist[j][k]

          # Each point needs to be moved away from or towards the other
          # point in proportion to how much error it has
          grad[k][0]+=((loc[k][0]-loc[j][0])/fakedist[j][k])*errorterm
          grad[k][1]+=((loc[k][1]-loc[j][1])/fakedist[j][k])*errorterm
```

e   length : 8561   lines : 281   Ln :

```python
        # Keep track of the total error
        totalerror+=abs(errorterm)
    print totalerror

    # If the answer got worse by moving the points, we are done
    if lasterror and lasterror<totalerror: break
    lasterror=totalerror

    # Move each of the points by the learning rate times the gradient
    for k in range(n):
      loc[k][0]-=rate*grad[k][0]
      loc[k][1]-=rate*grad[k][1]

  return loc

def draw2d(data,labels,jpeg='mds2d.jpg'):
  img=Image.new('RGB',(2000,2000),(255,255,255))
  draw=ImageDraw.Draw(img)
  for i in range(len(data)):
    x=(data[i][0]+0.5)*1000
    y=(data[i][1]+0.5)*1000
    draw.text((x,y),labels[i],(0,0,0))
  img.save(jpeg,'JPEG')
```
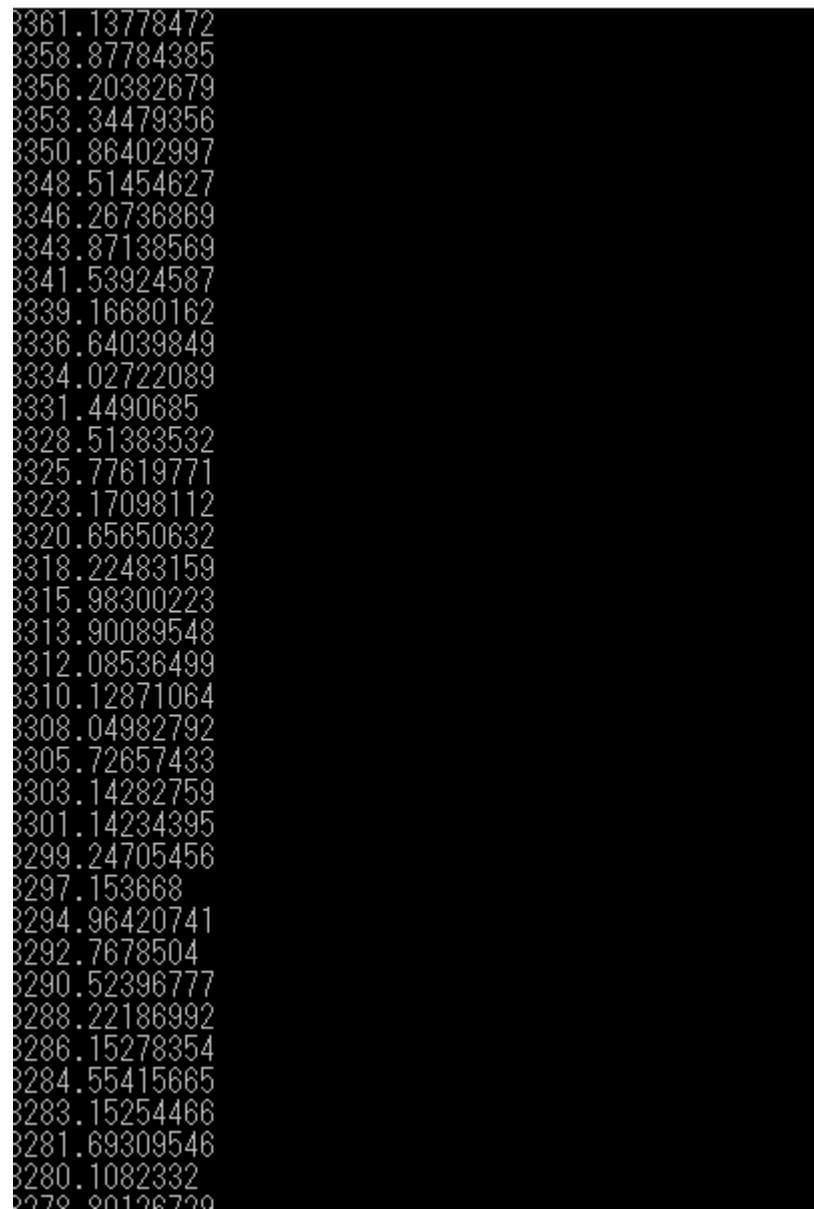
2.2 Results

There were 129 iterations and the graphic looks like this:

```
3361.13778472
3358.87784385
3356.20382679
3353.34479356
3350.86402997
3348.51454627
3346.26736869
3343.87138569
3341.53924587
3339.16680162
3336.64039849
3334.02722089
3331.4490685
3328.51383532
3325.77619771
3323.17098112
3320.65650632
3318.22483159
3315.98300223
3313.90089548
3312.08536499
3310.12871064
3308.04982792
3305.72657433
3303.14282759
3301.14234395
3299.24705456
3297.153668
3294.96420741
3292.7678504
3290.52396777
3288.22186992
3286.15278354
3284.55415665
3283.15254466
3281.69309546
3280.1082332
3278.80126739
```

New Hope Family Worship Center

A Van Berean Blog

Sunday Morning Bible Blog

Rev. B-B's random ramblings, really?

Rev. Miriam   Reverend Buf +

Would you like to Sing a Song?
jessdweil

House of Glory   Avare Nore Ore De   Unpluckable   Bread & Stones by the Wayray Family Faith in Cyberspace
ROSEKNOT COMMUNITY CHURCH   Referring Pastor   Welcome to Hope Abel's Blog
The First Sunday Thing   Mystical Ramblings

Valint   Joyful Peace

True Woman   Rhetoric and Civilization
Cee Harson Blog - Elevate Christian Network Ministry Outreach News   Seth's Line Upon Line
Love4od.LovePeople.

A Woman's Journey   Finding Grace Within   .... Altien Days ....   Francis X Archibald

Ditty Little Kisses   A Series of Links
Godaya Blog   Grinch's place
Life Long Learning Adventures

Hope for the Hormonal   leaning into life

The Family News Evening Spot   The Joy of Hope
My Journal, a homeschooling mom
Goddess Alchemy-A Priestess Journey   Suhi Suhi Vor this has gotta be the good life   On Arts International
Realities and Realizations
The Upside-Down Pastor's Wife   Wise Woman Alchemy

around the house   misadventures and whatnots   Gistcity Africa Media
Vivian Paul's Blog   ekh - e-Knowledge Hub

The Complicated Process of Understanding Love
CAROL WEEKS SPEAKS!   A Bucket of Grace

Drallery & Pathos
Extravandane   yeah
Until We Have Faces

Rena Girl | Relationships Life Love Enotions   Tru-Treasures   Upsellos

Anthology
In someone else's words   A Quiver Full   On the Carino
ENKLAVHE   Our little blue heaven
Kerenza Gray Jazz

My Little Thoughts and His Big Ones
News, Opinions and More   Soup to Nuts   T-Measure
How To Be and Live As an Ovarian
Adalia's Poetry   Henmira
The Little Things In A Simple Life   Dandelion Tea   This Walk   Latter-Day Liberty
Web Science and Digital Libraries Research Group   My life isn't perfect, but it does have perfect moments   Ada Robinson Illustration
c.o.n.s.t.a.n.c.e.
The Chaney scene   Jen & Carina Simpson   Synergies for Wise Living
Cougar Life   la Mama de Querida
The Blue Eyed Bunch   WELCOME TO FUNNY MICHAEL'S BLOG   Tanzanian Surgical Mission
The DeBlog: Duedee Tales   Adventures on Birch Creek Road   Karvamus Cebulla

My Path Is Thine - Elder Ka'ohu Ishibashi
Elder Adam Edwards In Paris
Welcome to j.s.price blog
Barcelona Dreamin

Elder Stringfellow: Serving the Lord in Milan Italy