Alexander Wilson
4/21/16
CS432 A9

1)Choose a blog or a newsfeed (or something similar with an Atom or RSS feed).

The Blog I chose was from a man up north living in nature (http://fishhookbay.blogspot.com). I am a big fan of nature and spend my summers in New Hampshire, Maine, and Vermont exploring. I chose this blog because he talks about his property and the day to day life while the seasons rotate. For this blog I chose the different seasons as my key words winter, fall, summer, and spring. I used the RSS feed from the blog

```
<?xml version='1.0' encoding='UTF-8'?><?xml-stylesheet href="http://www.blogger.com/styles/atom.css" type="text/css"?><feed
xmlns='http://www.w3.org/2005/Atom' xmlns:openSearch='http://a9.com/-/spec/opensearchrss/1.0/'
xmlns:blogger='http://schemas.google.com/blogger/2008' xmlns:georss='http://www.georss.org/georss'
xmlns:gd="http://schemas.google.com/g/2005" xmlns:thr='http://purl.org/syndication/thread/1.0'><id>tag:blogger.com,1999:blog-
1197930214366650364</id><updated>2016-04-17T18:05:27.735-07:00</updated><category term="late fall trees"/><category
term="things sure change fast"/><title type='text'>Sitting on a Dock in Our Bay</title><subtitle type='html'></subtitle><link
rel='http://schemas.google.com/g/2005#feed' type='application/atom+xml'
href='http://fishhookbay.blogspot.com/feeds/posts/default'/><link rel='self' type='application/atom+xml'
href='http://www.blogger.com/feeds/1197930214366650364/posts/default?max-results=1000'/><link rel='alternate' type='text/html'
href='http://fishhookbay.blogspot.com/'/><link rel='hub' href='http://pubsubhubbub.appspot.com/'/><author><name>Dr. Paul
Meartz</name><email>noreply@blogger.com</email><gd:image rel='http://schemas.google.com/g/2005#thumbnail' width='25'
height='32' src='http://3.bp.blogspot.com/_6mgtsAUYwsA/SrEhlJX7M6I/AAAAAAAAAAc/kclkcM_79B0/S220/pdm.gif'/></author><generator
version='7.00' uri='http://www.blogger.com'>Blogger</generator><openSearch:totalResults>100</openSearch:totalResults>
<openSearch:startIndex>1</openSearch:startIndex><openSearch:itemsPerPage>1000</openSearch:itemsPerPage><entry>
<id>tag:blogger.com,1999:blog-1197930214366650364.post-6825258509506583204</id><published>2016-04-17T18:05:00.001-
07:00</published><updated>2016-04-17T18:05:27.765-07:00</updated><title type='text'>Daisy on tour: Potato Lake</title><content
type='html'>&lt;blockquote&gt; &lt;p&gt;&amp;nbsp;&lt;/p&gt; &lt;p&gt;Daisy has been restless about checking out the boat
landings.&amp;nbsp; Those squirrels and chipmunks have been up to their evil, along with piles of other creatures.&amp;nbsp;
Potato Lake has a beautiful spot in the Northeast.&amp;nbsp; Looking down the lake to the East it almost has that mountain
```

Sample from→ (http://fishhookbay.blogspot.com/feeds/posts/default?max-results=100)

I opened the blog in Internet Explorer becuase it gave me the number of blog posts which was right at 100. After verifying that I was able to get my raw RSS data (Raw.txt).

2) I then went though and classified my RSS feed. I first separated each entry link into a .txt file. I used this list manually and automatically with fisherclassifer to classify each entry as a different season. I found docclass.py online which had the fisherclassifier subclass in it. It was able to

parse the RSS feed to eliminate unnecessary HTML. After I classified the first 50 parsed links the program q2.py predicited the rest.

1	Title	Predicted	Actual	cprob()
2	daisy on tour potato lake	winter	winter	0.9
3	ice out ii	winter	winter	0.9
4	and now lake ice 2	winter	winter	0.9
5	and now lake ice 2	winter	winter	0.9
6	white snow again	winter	winter	0.9
7	them deer	winter	winter	0.9
8	spring rains	winter	spring	0.3
9	early spring	spring	spring	0.9
10	snow and more cold	spring	spring	0.9
11	snow as dust period	spring	winter	0.2
12	winter for duration	winter	winter	0.8
13	ice	winter	winter	0.9
14	snow snow snow	winter	winter	0.9
15	snow here to stay	winter	winter	0.9
16	snow sort of	winter	winter	0.9
17	a darker day	spring	winter	0.1
18	swans passing by	spring	winter	0.4
19	warm water	spring	spring	0.9
20	animal fest	spring	summer	0.3
21	spring	spring	spring	0.9
22	maybe spring	spring	spring	0.9
23	cold	winter	winter	0.9
24	looks like spring	spring	spring	0.9
25	snow	winter	winter	0.9
26	brrrr	winter	winter	0.8
27	sun	spring	winter	0.3
28	a quite christmas	winter	winter	0.9
29	indecision	fall	winter	0.4
30	ice noise	winter	winter	0.9
31	ice over	winter	winter	0.9
32	winter for duration	winter	winter	0.9
33	winter nears	winter	winter	0.9
34	swimming ends	winter	fall	0.1
35	squirrels planning bad winter	winter	fall	0.7
36	fall ahead	fall	fall	0.8
37	fall watch	fall	fall	0.8
38	them loons now them geese	summer	fall	0.8

3) To calculate the precision, recall, and F- measure I used the python. I first had to find the number of true positives, false negatives, and false positives. The Fall element seemed to be the most accurate while the winter fell in a close second. I used the number of TP, FP, and FNs to calculate the precision recall and F-Measure. I think spring had the most difficulty because it had the least amount of data to base its training off.

	precision [tp/(tp+fp)	recall [tp/(tp+fn)]	F-Measure 2*P*R/(P+R)
winter	30/45	0.666666667	0.678945
summer	13/20	0.65456	0.7067
spring	7/16.	0.437525	0.54521
fall	14/19	0.736842105	0.8864