**Student Name: Ekal Golas**

NetID: exg140230

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

University of Texas at Dallas

CS 6322.001 : Information Retrieval

Spring 2016

Take Home Quiz # 3

Instructor: Dr. Sanda Harabagiu

Grader: Ramon Maldonado

Issued: April 18th 2016 Due April 20th 2016 in class

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Query Expansion QUIZ**

Use Automatic Local Analysis based on Metric Clusters to expand the following query:

*Original QUERY*:**earthquake Ecuador**

– When the local collection has 4 documents (N=4) *Vl=??? Sl=???*

**Document 1:** One person was dead after a magnitude-7.8 earthquake occurredSaturday evening on the coast of Ecuador.

**Document 2:** An earthquake with magnitude 7.4 occurred near Esmeraldas, Ecuadorat 23:58:37.40 UTC on Apr 16, 2016.

**Document 3:** A 7.8-magnitude earthquake struck near Ecuador's coast Saturday,shaking homes 100 miles away in the capital of Quito and leaving 28 people dead.

**Document 4:** A powerful 7.8 magnitude earthquake shook Ecuador's central coastSaturday, cracking buildings and rattling homes as far away as the capital of Quito.

– Find what is the local vocabulary *Vl=???* as well as the local stems *Sl=???*

Then, compute the distance between stems, given that *the distance r(ki, kj) between* *two stem keywords ki and kj is given by the number of words between them in the same document. If ki and kj are in distinct documents we take r(k i, kj)=.* This allows you tocompute the correlations terms, using the formula:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  | *cu* ,*v*  |  |  | 1 | |  |  |  |  |  |
|  | *r*(*k* | , *k* |  | ) |  |  |  |
|  |  | *ki* *V* ( *su* ) *k j* *V* ( *su* ) | | *j* |  |  |  |
|  |  | *i* |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

The correlation values between the stems from the local collection inform a correlation matrix from which you can infer the metric clusters for the keywords “earthquake” and “Ecuador”. You are required to generate the clusters of size 3, i.e. find the clusters *Su(n)* *defined as local metric clusters around each stem su from the original query, as we assume n=3 for each metric cluster!!!*

Show how you have built the metric clusters for keywords “earthquake” and “Ecuador” and show the expanded query that they determine.

In addition, generate the normalized metric clusters and the resulting expanded query when you consider the normalization formula:

*su*,*v*  *c**u*,*v*

| *V* (*su* ) | | *V* (*sv* ) |

## Answer

Vi: Relevant words (Ignoring the stop words, special characters and numbers but count them in the distance):

Size: 32

Matrix: (Word={document number = position in that document})

person={1=2} rattling={4=14} one={1=1} earthquake={1=8, 2=2, 3=3, 4=5} occurred={1=9, 2=6} shaking={3=9} people={3=22} capital={3=16, 4=21} coast={1=14, 3=7, 4=9} struck={3=4} leaving={3=20} after={1=5} esmeraldas={2=8} ecuadors={3=6, 4=7} buildings={4=12} saturday={1=10, 3=8, 4=10} dead={1=4, 3=23} homes={3=10, 4=15} central={4=8} evening={1=11} quito={3=18, 4=23} apr={2=14} ecuador={1=16, 2=9} magnitude={1=7, 2=4, 3=2, 4=4} utc={2=12} far={4=17} miles={3=12} powerful={4=2} away={3=13, 4=18} near={2=7, 3=5} cracking={4=11} shook={4=6}

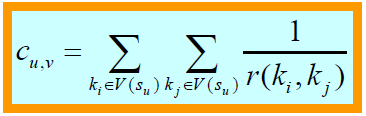
Si: After stemming the vocabulary

Size: 30

Matrix: (Stem=[tokens for that stem])

person=[person] rattling=[rattling] one=[one] shake=[shook, shaking] earthquake=[earthquake] occurred=[occurred] people=[people] capital=[capital] coast=[coast] struck=[struck] leaving=[leaving] after=[after] esmeraldas=[esmeraldas] buildings=[buildings] saturday=[saturday] dead=[dead] homes=[homes] central=[central] evening=[evening] quito=[quito] apr=[apr] ecuador=[ecuador, ecuadors] utc=[utc] magnitude=[magnitude] far=[far] miles=[miles] powerful=[powerful] away=[away] near=[near] cracking=[cracking]

Using the formula:



Metric Clusters matrix: (u v : Cu,v) separated by commas

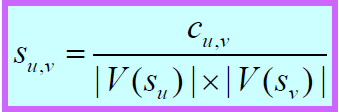
one person : 1.0, shake rattling : 0.125, earthquake person : 0.16666666666666666, earthquake rattling : 0.1111111111111111, earthquake one : 0.14285714285714285, earthquake shake : 1.1666666666666667, occurred person : 0.14285714285714285, occurred one : 0.125, occurred earthquake : 1.25, people shake : 0.07692307692307693, people earthquake : 0.05263157894736842, capital rattling : 0.14285714285714285, capital shake : 0.2095238095238095, capital earthquake : 0.13942307692307693, capital people : 0.16666666666666666, coast person : 0.08333333333333333, coast rattling : 0.2, coast one : 0.07692307692307693, coast shake : 0.8333333333333333, coast earthquake : 0.6666666666666666, coast occurred : 0.2, coast people : 0.06666666666666667, coast capital : 0.19444444444444442, struck shake : 0.2, struck earthquake : 1.0, struck people : 0.05555555555555555, struck capital : 0.08333333333333333, struck coast : 0.3333333333333333, leaving shake : 0.09090909090909091, leaving earthquake : 0.058823529411764705, leaving people : 0.5, leaving capital : 0.25, leaving coast : 0.07692307692307693, leaving struck : 0.0625, after person : 0.3333333333333333, after one : 0.25, after earthquake : 0.3333333333333333, after occurred : 0.25, after coast : 0.1111111111111111, esmeraldas earthquake : 0.16666666666666666, esmeraldas occurred : 0.5, buildings rattling : 0.5, buildings shake : 0.16666666666666666, buildings earthquake : 0.14285714285714285, buildings capital : 0.1111111111111111, buildings coast : 0.3333333333333333, saturday person : 0.125, saturday rattling : 0.25, saturday one : 0.1111111111111111, saturday shake : 1.25, saturday earthquake : 0.8999999999999999, saturday occurred : 1.0, saturday people : 0.07142857142857142, saturday capital : 0.2159090909090909, saturday coast : 2.25, saturday struck : 0.25, saturday leaving : 0.08333333333333333, saturday after : 0.2, saturday buildings : 0.5, dead person : 0.5, dead one : 0.3333333333333333, dead shake : 0.07142857142857142, dead earthquake : 0.3, dead occurred : 0.2, dead people : 1.0, dead capital : 0.14285714285714285, dead coast : 0.1625, dead struck : 0.05263157894736842, dead leaving : 0.3333333333333333, dead after : 1.0, dead saturday : 0.23333333333333334, homes rattling : 1.0, homes shake : 1.1111111111111112, homes earthquake : 0.24285714285714285, homes people : 0.08333333333333333, homes capital : 0.3333333333333333, homes coast : 0.5, homes struck : 0.16666666666666666, homes leaving : 0.1, homes buildings : 0.3333333333333333, homes saturday : 0.7, homes dead : 0.07692307692307693, central rattling : 0.16666666666666666, central shake : 0.5, central earthquake : 0.3333333333333333, central capital : 0.07692307692307693, central coast : 1.0, central buildings : 0.25, central saturday : 0.5, central homes : 0.14285714285714285, evening person : 0.1111111111111111, evening one : 0.1, evening earthquake : 0.3333333333333333, evening occurred : 0.5, evening coast : 0.3333333333333333, evening after : 0.16666666666666666, evening saturday : 1.0, evening dead : 0.14285714285714285, quito rattling : 0.1111111111111111, quito shake : 0.16993464052287582, quito earthquake : 0.12222222222222222, quito people : 0.25, quito capital : 1.0, quito coast : 0.16233766233766234, quito struck : 0.07142857142857142, quito leaving : 0.5, quito buildings : 0.09090909090909091, quito saturday : 0.17692307692307693, quito dead : 0.2, quito homes : 0.25, quito central : 0.06666666666666667, apr earthquake : 0.08333333333333333, apr occurred : 0.125, apr esmeraldas : 0.16666666666666666, ecuador person : 0.07142857142857142, ecuador rattling : 0.14285714285714285, ecuador one : 0.06666666666666667, ecuador shake : 1.3333333333333333, ecuador earthquake : 1.1011904761904763, ecuador occurred : 0.47619047619047616, ecuador people : 0.0625, ecuador capital : 0.17142857142857143, ecuador coast : 2.0, ecuador struck : 0.5, ecuador leaving : 0.07142857142857142, ecuador after : 0.09090909090909091, ecuador esmeraldas : 1.0, ecuador buildings : 0.2, ecuador saturday : 1.0, ecuador dead : 0.14215686274509803, ecuador homes : 0.375, ecuador central : 1.0, ecuador evening : 0.2, ecuador quito : 0.14583333333333331, ecuador apr : 0.2, utc earthquake : 0.1, utc occurred : 0.16666666666666666, utc esmeraldas : 0.25, utc apr : 0.5, utc ecuador : 0.3333333333333333, magnitude person : 0.2, magnitude rattling : 0.1, magnitude one : 0.16666666666666666, magnitude shake : 0.6428571428571428, magnitude earthquake : 3.5, magnitude occurred : 1.0, magnitude people : 0.05, magnitude capital : 0.13025210084033612, magnitude coast : 0.5428571428571429, magnitude struck : 0.5, magnitude leaving : 0.05555555555555555, magnitude after : 0.5, magnitude esmeraldas : 0.25, magnitude buildings : 0.125, magnitude saturday : 0.6666666666666666, magnitude dead : 0.38095238095238093, magnitude homes : 0.2159090909090909, magnitude central : 0.25, magnitude evening : 0.25, magnitude quito : 0.11513157894736842, magnitude apr : 0.1, magnitude ecuador : 0.8944444444444444, magnitude utc : 0.125, far rattling : 0.3333333333333333, far shake : 0.09090909090909091, far earthquake : 0.08333333333333333, far capital : 0.25, far coast : 0.125, far buildings : 0.2, far saturday : 0.14285714285714285, far homes : 0.5, far central : 0.1111111111111111, far quito : 0.16666666666666666, far ecuador : 0.1, far magnitude : 0.07692307692307693, miles shake : 0.3333333333333333, miles earthquake : 0.1111111111111111, miles people : 0.1, miles capital : 0.25, miles coast : 0.2, miles struck : 0.125, miles leaving : 0.125, miles saturday : 0.25, miles dead : 0.09090909090909091, miles homes : 0.5, miles quito : 0.16666666666666666, miles ecuador : 0.16666666666666666, miles magnitude : 0.1, powerful rattling : 0.08333333333333333, powerful shake : 0.25, powerful earthquake : 0.3333333333333333, powerful capital : 0.05263157894736842, powerful coast : 0.14285714285714285, powerful buildings : 0.1, powerful saturday : 0.125, powerful homes : 0.07692307692307693, powerful central : 0.16666666666666666, powerful quito : 0.047619047619047616, powerful ecuador : 0.2, powerful magnitude : 0.5, powerful far : 0.06666666666666667, away rattling : 0.25, away shake : 0.3333333333333333, away earthquake : 0.17692307692307693, away people : 0.1111111111111111, away capital : 0.6666666666666666, away coast : 0.2777777777777778, away struck : 0.1111111111111111, away leaving : 0.14285714285714285, away buildings : 0.16666666666666666, away saturday : 0.325, away dead : 0.1, away homes : 0.6666666666666666, away central : 0.1, away quito : 0.4, away ecuador : 0.23376623376623376, away magnitude : 0.16233766233766234, away far : 1.0, away miles : 1.0, away powerful : 0.0625, near shake : 0.25, near earthquake : 0.7, near occurred : 1.0, near people : 0.058823529411764705, near capital : 0.09090909090909091, near coast : 0.5, near struck : 1.0, near leaving : 0.06666666666666667, near esmeraldas : 1.0, near saturday : 0.3333333333333333, near dead : 0.05555555555555555, near homes : 0.2, near quito : 0.07692307692307693, near apr : 0.14285714285714285, near ecuador : 1.5, near utc : 0.2, near magnitude : 0.6666666666666666, near miles : 0.14285714285714285, near away : 0.125, cracking rattling : 0.3333333333333333, cracking shake : 0.2, cracking earthquake : 0.16666666666666666, cracking capital : 0.1, cracking coast : 0.5, cracking buildings : 1.0, cracking saturday : 1.0, cracking homes : 0.25, cracking central : 0.3333333333333333, cracking quito : 0.08333333333333333, cracking ecuador : 0.25, cracking magnitude : 0.14285714285714285, cracking far : 0.16666666666666666, cracking powerful : 0.1111111111111111, cracking away : 0.14285714285714285,

Top 3 terms for query expansion:

earthquake shake : 1.1666666666666667 earthquake occurred : 1.25 earthquake magnitude : 3.5

ecuador shake : 1.3333333333333333 ecuador near : 1.5 ecuador coast : 2.0

Using the formula:



Metric Clusters after normalization: (u v : Cu,v) separated by commas

one person : 1.0, shake rattling : 0.0625, earthquake person : 0.16666666666666666, earthquake rattling : 0.1111111111111111, earthquake one : 0.14285714285714285, earthquake shake : 0.5833333333333334, occurred person : 0.14285714285714285, occurred one : 0.125, occurred earthquake : 1.25, people shake : 0.038461538461538464, people earthquake : 0.05263157894736842, capital rattling : 0.14285714285714285, capital shake : 0.10476190476190475, capital earthquake : 0.13942307692307693, capital people : 0.16666666666666666, coast person : 0.08333333333333333, coast rattling : 0.2, coast one : 0.07692307692307693, coast shake : 0.41666666666666663, coast earthquake : 0.6666666666666666, coast occurred : 0.2, coast people : 0.06666666666666667, coast capital : 0.19444444444444442, struck shake : 0.1, struck earthquake : 1.0, struck people : 0.05555555555555555, struck capital : 0.08333333333333333, struck coast : 0.3333333333333333, leaving shake : 0.045454545454545456, leaving earthquake : 0.058823529411764705, leaving people : 0.5, leaving capital : 0.25, leaving coast : 0.07692307692307693, leaving struck : 0.0625, after person : 0.3333333333333333, after one : 0.25, after earthquake : 0.3333333333333333, after occurred : 0.25, after coast : 0.1111111111111111, esmeraldas earthquake : 0.16666666666666666, esmeraldas occurred : 0.5, buildings rattling : 0.5, buildings shake : 0.08333333333333333, buildings earthquake : 0.14285714285714285, buildings capital : 0.1111111111111111, buildings coast : 0.3333333333333333, saturday person : 0.125, saturday rattling : 0.25, saturday one : 0.1111111111111111, saturday shake : 0.625, saturday earthquake : 0.8999999999999999, saturday occurred : 1.0, saturday people : 0.07142857142857142, saturday capital : 0.2159090909090909, saturday coast : 2.25, saturday struck : 0.25, saturday leaving : 0.08333333333333333, saturday after : 0.2, saturday buildings : 0.5, dead person : 0.5, dead one : 0.3333333333333333, dead shake : 0.03571428571428571, dead earthquake : 0.3, dead occurred : 0.2, dead people : 1.0, dead capital : 0.14285714285714285, dead coast : 0.1625, dead struck : 0.05263157894736842, dead leaving : 0.3333333333333333, dead after : 1.0, dead saturday : 0.23333333333333334, homes rattling : 1.0, homes shake : 0.5555555555555556, homes earthquake : 0.24285714285714285, homes people : 0.08333333333333333, homes capital : 0.3333333333333333, homes coast : 0.5, homes struck : 0.16666666666666666, homes leaving : 0.1, homes buildings : 0.3333333333333333, homes saturday : 0.7, homes dead : 0.07692307692307693, central rattling : 0.16666666666666666, central shake : 0.25, central earthquake : 0.3333333333333333, central capital : 0.07692307692307693, central coast : 1.0, central buildings : 0.25, central saturday : 0.5, central homes : 0.14285714285714285, evening person : 0.1111111111111111, evening one : 0.1, evening earthquake : 0.3333333333333333, evening occurred : 0.5, evening coast : 0.3333333333333333, evening after : 0.16666666666666666, evening saturday : 1.0, evening dead : 0.14285714285714285, quito rattling : 0.1111111111111111, quito shake : 0.08496732026143791, quito earthquake : 0.12222222222222222, quito people : 0.25, quito capital : 1.0, quito coast : 0.16233766233766234, quito struck : 0.07142857142857142, quito leaving : 0.5, quito buildings : 0.09090909090909091, quito saturday : 0.17692307692307693, quito dead : 0.2, quito homes : 0.25, quito central : 0.06666666666666667, apr earthquake : 0.08333333333333333, apr occurred : 0.125, apr esmeraldas : 0.16666666666666666, ecuador person : 0.03571428571428571, ecuador rattling : 0.07142857142857142, ecuador one : 0.03333333333333333, ecuador shake : 0.3333333333333333, ecuador earthquake : 0.5505952380952381, ecuador occurred : 0.23809523809523808, ecuador people : 0.03125, ecuador capital : 0.08571428571428572, ecuador coast : 1.0, ecuador struck : 0.25, ecuador leaving : 0.03571428571428571, ecuador after : 0.045454545454545456, ecuador esmeraldas : 0.5, ecuador buildings : 0.1, ecuador saturday : 0.5, ecuador dead : 0.07107843137254902, ecuador homes : 0.1875, ecuador central : 0.5, ecuador evening : 0.1, ecuador quito : 0.07291666666666666, ecuador apr : 0.1, utc earthquake : 0.1, utc occurred : 0.16666666666666666, utc esmeraldas : 0.25, utc apr : 0.5, utc ecuador : 0.16666666666666666, magnitude person : 0.2, magnitude rattling : 0.1, magnitude one : 0.16666666666666666, magnitude shake : 0.3214285714285714, magnitude earthquake : 3.5, magnitude occurred : 1.0, magnitude people : 0.05, magnitude capital : 0.13025210084033612, magnitude coast : 0.5428571428571429, magnitude struck : 0.5, magnitude leaving : 0.05555555555555555, magnitude after : 0.5, magnitude esmeraldas : 0.25, magnitude buildings : 0.125, magnitude saturday : 0.6666666666666666, magnitude dead : 0.38095238095238093, magnitude homes : 0.2159090909090909, magnitude central : 0.25, magnitude evening : 0.25, magnitude quito : 0.11513157894736842, magnitude apr : 0.1, magnitude ecuador : 0.4472222222222222, magnitude utc : 0.125, far rattling : 0.3333333333333333, far shake : 0.045454545454545456, far earthquake : 0.08333333333333333, far capital : 0.25, far coast : 0.125, far buildings : 0.2, far saturday : 0.14285714285714285, far homes : 0.5, far central : 0.1111111111111111, far quito : 0.16666666666666666, far ecuador : 0.05, far magnitude : 0.07692307692307693, miles shake : 0.16666666666666666, miles earthquake : 0.1111111111111111, miles people : 0.1, miles capital : 0.25, miles coast : 0.2, miles struck : 0.125, miles leaving : 0.125, miles saturday : 0.25, miles dead : 0.09090909090909091, miles homes : 0.5, miles quito : 0.16666666666666666, miles ecuador : 0.08333333333333333, miles magnitude : 0.1, powerful rattling : 0.08333333333333333, powerful shake : 0.125, powerful earthquake : 0.3333333333333333, powerful capital : 0.05263157894736842, powerful coast : 0.14285714285714285, powerful buildings : 0.1, powerful saturday : 0.125, powerful homes : 0.07692307692307693, powerful central : 0.16666666666666666, powerful quito : 0.047619047619047616, powerful ecuador : 0.1, powerful magnitude : 0.5, powerful far : 0.06666666666666667, away rattling : 0.25, away shake : 0.16666666666666666, away earthquake : 0.17692307692307693, away people : 0.1111111111111111, away capital : 0.6666666666666666, away coast : 0.2777777777777778, away struck : 0.1111111111111111, away leaving : 0.14285714285714285, away buildings : 0.16666666666666666, away saturday : 0.325, away dead : 0.1, away homes : 0.6666666666666666, away central : 0.1, away quito : 0.4, away ecuador : 0.11688311688311688, away magnitude : 0.16233766233766234, away far : 1.0, away miles : 1.0, away powerful : 0.0625, near shake : 0.125, near earthquake : 0.7, near occurred : 1.0, near people : 0.058823529411764705, near capital : 0.09090909090909091, near coast : 0.5, near struck : 1.0, near leaving : 0.06666666666666667, near esmeraldas : 1.0, near saturday : 0.3333333333333333, near dead : 0.05555555555555555, near homes : 0.2, near quito : 0.07692307692307693, near apr : 0.14285714285714285, near ecuador : 0.75, near utc : 0.2, near magnitude : 0.6666666666666666, near miles : 0.14285714285714285, near away : 0.125, cracking rattling : 0.3333333333333333, cracking shake : 0.1, cracking earthquake : 0.16666666666666666, cracking capital : 0.1, cracking coast : 0.5, cracking buildings : 1.0, cracking saturday : 1.0, cracking homes : 0.25, cracking central : 0.3333333333333333, cracking quito : 0.08333333333333333, cracking ecuador : 0.125, cracking magnitude : 0.14285714285714285, cracking far : 0.16666666666666666, cracking powerful : 0.1111111111111111, cracking away : 0.14285714285714285,

Top 3 terms for query expansion after normalization:

earthquake struck : 1.0 earthquake magnitude : 3.5 earthquake occurred : 1.25

ecuador earthquake : 0.5505952380952381 ecuador near : 0.75 ecuador coast : 1.0