

Final Project for Course CS 544 Natural Language Processing

To: Professor Eduard Hovy

By: Elnaz Nouri 32 70 46 88 63

1) Information on Data:

The text data for this assignment was downloaded from the news articles on [BBC website](#).

Doc #	Doc Topic	Number of Subsections	Number of Lines	Number of lines in ideal summary
1	gay rights in brazil	1	13	4
2	Einstein star	3	36	11
3	Libya: rebels	3	27	8
4	pregnancy	1	17	5
5	royal weds	1	13	4
6	General Motors	1	10	3
7	al pacino	1	10	3
8	Obama	1	37	9
9	Georgia eradication	1	39	11
10	elephants	4	43	12
11	hack	4	37	10
12	actress dies	1	12	4
13	McCartney marry	1	7	2

As mentioned in the project description the first 4 texts (shown in grey in the table above) were used as test docs and the 6 next documents were used as development. There are 3 extras test files as well.

The input – output – trace – query of running the system on the two (yellow highlighted) texts are given in this document. [For the output and results of the rest of the documents go to the files included in the folder.]

2) Topic Scoring Modules Used in the program are:

i. Scoring method: Word frequency (tf, tf-idf)

For all the words in the document the tf- idf values are measured.

```
private static TreeMap<String, Double> term_tdf( TreeMap<String,
Integer> map, int doc_size_1) {

    TreeMap<String, Double> new_map_tf=new TreeMap<String, Double>();
    Set keys=map.keySet();
    for(Iterator i=keys.iterator();i.hasNext();){
        String key= (String) i.next();
        double tf= (double) ( (double) map.get(key) )/
(double) (doc_size_1);
        new_map_tf.put(key, tf);
    }
    return new_map_tf;
}

private static TreeMap<String, Double> term_idf( TreeMap<String, Integer>
map, ArrayList<TreeMap<String, Integer>> all) {

    TreeMap<String, Double> new_map_idf=new TreeMap<String,
Double>();
    Set keys=map.keySet();
    for(Iterator i=keys.iterator();i.hasNext();){
        String key= (String) i.next();
        double idf=1;
        for (int j=0;j<all.size();j++){
            if(all.get(j).containsKey(key)){
                idf=idf+1;
            }
        }
        double idf_value=Math.log(((double) (all.size()))/idf);
        new_map_idf.put(key, idf_value);
    }
    return new_map_idf;
}

private static TreeMap<String, Double> term_tdf_idf( TreeMap<String, Double>
word_tdf_1, TreeMap<String, Double> word_idf_1) {

    if(word_tdf_1.size() != word_idf_1.size()){
        System.out.println("This is a problem now!");
        System.exit(1);
    }
    TreeMap<String, Double> new_map_tf_idf=new TreeMap<String,
Double>();
    Set keys=word_tdf_1.keySet();
    for(Iterator i=keys.iterator();i.hasNext();){
        String key= (String) i.next();
        double tf_idf=-1;
        tf_idf= word_tdf_1.get(key)*word_idf_1.get(key);
        new_map_tf_idf.put(key, tf_idf);
    }
    return new_map_tf_idf;
}
```

After calculation of all tf-idf values for all the words in the sentences, I calculate a vector for each sentence, the vector contains an entry for each of the words in the current document, if the sentence contains any of the words the vector entry is set to be the tf-idf value for that word, and if not, it is set to be zero.

I also come up with another score for each sentence, in which the vector is normalized and then the values of the vector entries are summed up to make up a new value for the vector score.

```
private static ArrayList<Double> calculate_sentence_score(
ArrayList<ArrayList<Double>> sentence_vectors) {

    ArrayList<Double> scores= new ArrayList<Double>();
    for(int
sentence_index=0;sentence_index<sentence_vectors.size();sentence_index++){
        ArrayList<Double>
sentence_normal=sentence_vectors.get(sentence_index);
        double temp=0;
        for(int i=0;i<sentence_normal.size();i++){
            temp=temp+sentence_normal.get(i);
        }
        scores.add(temp);
    }
    return scores;
}
```

ii. Position

The higher a sentence is in a document the higher the score it's assigned to. Also sentences in the beginning of the paragraphs and after the headlines are assigned higher scores.

```
private static ArrayList<Double> readPosition(String inputfile,
ArrayList<Double> headerlines) throws IOException {
    //System.out.println("Come here 4");
    FileInputStream fstream = new FileInputStream(inputfile);
    DataInputStream in = new DataInputStream(fstream);
    BufferedReader br = new BufferedReader(new InputStreamReader(in));
    ArrayList<Double> poslist=new ArrayList<Double>();
    String strLine;
    int position=0;
    while ((strLine = (br.readLine()) ) != null) {
        position++;
        if(headerlines.get(position-1)>1){
            poslist.add((double)headerlines.size()-position+5);
        }else{
            poslist.add((double)headerlines.size()-position);
        }
    }
    return poslist;
}
```

iii. Headings

The input text files have html tags that show whether they are headers or the beginning of the sections and paragraphs.

```
public static ArrayList<Double> readFileHeader(String rulefile) throws
IOException{
    //System.out.println("Come here 2");
    FileInputStream fstream = new FileInputStream(rulefile);
    DataInputStream in = new DataInputStream(fstream);
    BufferedReader br = new BufferedReader(new InputStreamReader(in));
    ArrayList<Double> headerlist=new ArrayList<Double>();
    String strLine;
    while ((strLine = (br.readLine()) ) != null)    {
        strLine=strLine.toLowerCase();
        if(strLine.contains("<h>")){
            headerlist.add(10.0);
        }else if(strLine.contains("<h1>")){
            headerlist.add(5.0);
        }else if(strLine.contains("<h2>")){
            headerlist.add(4.0);
        }else if(strLine.contains("<h3>")){
            headerlist.add(3.0);
        }else if(strLine.contains("<p>")){
            headerlist.add(2.0);
        }else{
            headerlist.add(0.0);
        }
    }
    return headerlist;
}
```

iv. Que Phrases

In this method, I look for words or phrases that indicate centrality.

```
private static ArrayList<Double> readCue(String inputfile) throws IOException
{
    //System.out.println("Come here 6");
    FileInputStream fstream = new FileInputStream(inputfile);
    DataInputStream in = new DataInputStream(fstream);
    BufferedReader br = new BufferedReader(new InputStreamReader(in));
    ArrayList<Double> cuelist=new ArrayList<Double>();
    String strLine;
    while ((strLine = (br.readLine()) ) != null)    {
        strLine=strLine.toLowerCase();
        if(strLine.contains("important")){
            cuelist.add(4.0);
        }else if(strLine.contains("significant")){
            cuelist.add(2.0);
        }else if(strLine.contains("conclusion") || strLine.contains("sum
up") || strLine.contains("finally") || strLine.contains("because") ){
            cuelist.add(3.0);
        }else if(strLine.contains("however") ||
strLine.contains("therefore") || strLine.contains("but") ||
strLine.contains("although")){
```

```

        cuelist.add(1.0);
    }else{
        cuelist.add(0.0);
    }
}

return cuelist;
}

```

v. Stigma Phrases

In this method I look for phrases that indicated the sentences after them, might not be as central to the paragraph and I score them lower. I look for the verb say and it's derivatives and quotations marks and conjunctions.

```

private static ArrayList<Double> readStigma(String inputfile) throws
IOException {
    //System.out.println("Come here 3");
    FileInputStream fstream = new FileInputStream(inputfile);
    DataInputStream in = new DataInputStream(fstream);
    BufferedReader br = new BufferedReader(new
InputStreamReader(in));
    ArrayList<Double> stigmalist=new ArrayList<Double>();
    String strLine;
    while ((strLine = (br.readLine()) ) != null) {
        strLine=strLine.toLowerCase();
        if(strLine.contains("but") || strLine.contains("example")){
            stigmalist.add(1.0);
        }else if(strLine.contains("\"")){
            stigmalist.add(2.0);
        }else if(strLine.contains("say") ||
strLine.contains("said")|| strLine.contains("says") ||
strLine.contains("announces") ){
            stigmalist.add(3.0);
        }else if(strLine.contains("however") ||
strLine.contains("therefore")|| strLine.contains("but") ||
strLine.contains("although")){
            stigmalist.add(1.0);
        }else{
            stigmalist.add(10.0);
        }
    }
    return stigmalist;
}

```

vi. Lexical Chain

In this method, I look for the dependencies between the sentences. A 2 dimensional graph(number of sentences* number of sentences) is made that shows which of the sentences share certain words.

I came to observe three problems with my lexical chain module,

first, it is not reasonable to look for all the words, certain words should be selected as candidates, which are the words that most likely represent the content of the document and therefore have high tf-idf scores.

Second, a word in a sentence might need to get connected to its synonyms in other sentences, therefore I look up the word in wordnet and use the synset of the word for making the comparison and building the edges of the graph.

Third, longer sentences were given higher score by this method, because they included many of the main words of the document, but that did not necessarily mean that they were the best sentences. For example in document one:

the ruling makes brazil one of very few south american nations, after argentina and urug"uay, to allow gay unions with benefits similar to those afforded a heterosexual married couple. " was always chosen.

Note: it is also possible to make this graph based on the words in the query. By marking out the sentences that contain the specific words that the query contains.

```
private static ArrayList<Double> score_module_Lexical_Chain(ArrayList<String>
inputlines, ArrayList<String> sorted_words, ArrayList<Integer> length_12) {

    ArrayList<Double> score_2= new ArrayList<Double>();

    System.setProperty("wordnet.database.dir", "C:\\\\Program Files
(x86)\\\\WordNet\\\\2.1\\\\dict\\\\");
    WordNetDatabase database = WordNetDatabase.getFileInstance();

    System.out.println("Number of sentences: "+inputlines.size());
    System.out.println("sorted_words: "+sorted_words.size());

    ArrayList<ArrayList<String>> sentence_words_synset= new
ArrayList<ArrayList<String>>();

    for(int i=0;i<sorted_words.size();i++){
        ArrayList<String> current_word_set= new
ArrayList<String>();
        String wordForm = sorted_words.get(i);
        current_word_set.add(wordForm);
        Synset[] synsets = database.getSynsets(wordForm);
        //System.out.println("* current_word_setk :
"+current_word_set.size());
        //System.out.println(current_word_set.toString());
        if (synsets.length > 0){
            //System.out.println("The following synsets contain
'" + wordForm + "' or a possible base form of that text:");
            for (int k = 0; k < synsets.length; k++)
            {
                String[] wordForms =
synsets[k].getWordForms();
                for (int j = 0; j < wordForms.length;
j++)
                {
```

```

//System.out.print( wordForms[j] +
"\t");
        if( !
current_word_set.contains(wordForms[j]) ){
            current_word_set.add(wordForms[j]);
        }
    }
    //System.out.println();
}

}
sentence_words_synset.add(current_word_set);
System.out.println(current_word_set.toString());

}
System.out.println("* * * Just to check :
"+sentence_words_synset.size());

int[][] connectivity=new
int[inputlines.size()][inputlines.size()];
for(int i=0;i<inputlines.size();i++){
    for(int j=0;j<inputlines.size();j++){
        connectivity[i][j]=0;
    }
}
for(int i=0;i<sentence_words_synset.size();i++){
    for(int j=0;j<inputlines.size();j++){
        for(int h=j;h<inputlines.size();h++){
            if(j != h){
                boolean found_j=false;
                boolean found_h=false;

                for(int k=0;
k<sentence_words_synset.get(i).size();k++){
                    String
word=sentence_words_synset.get(i).get(k);

                    if(inputlines.get(j).contains(word)){
                        found_j=true;
                    }

                    if(inputlines.get(h).contains(word)){
                        found_h=true;
                    }
                    if(found_h && found_j){

                        //System.out.println(sentence_words_synset.get(i));
                        System.out.println(word
+"\\tin both "+j +"\\t"+h +"\\t");

                        connectivity[h][j]=connectivity[h][j]+1;

                        connectivity[j][h]=connectivity[j][h]+1;

                        break;

```

```

    }
    }
    }
    }

String output="";
for(int j=0;j<inputlines.size();j++){
    double temp=0;
    for(int h=0;h<inputlines.size();h++){
        temp=temp+connectivity[j][h];
        output=output+"\t"+connectivity[j][h];
    }
    score_2.add( temp);
    output=output+"\t="+temp+"\t"+temp/length_12.get(j)+"\n\n";
}
System.out.println(output);

return score_2;
}

```

vii. Sentence Vector Similarity to Query Vector Similarity

The sentences and the query are all made into vectors, the vectors are normalized and the dot product of them is calculated, the higher the $\cos\theta$ the higher similarity between the two.

```

static ArrayList<Double> make_query_vector(ArrayList<String> sorted_words,
ArrayList<Double> sorted_scores, ArrayList<String> inputlines){
    ArrayList<Double> query_vector= new ArrayList<Double>();
    for(int term=0;term<sorted_words.size();term++){
        if(query_vector.contains(sorted_words.get(term))){
            query_vector.add(sorted_scores.get(term));
        }else{
            query_vector.add(0.0);
        }
    }
    return query_vector;
}

private static ArrayList<Double> calculate_similarity_score(
ArrayList<ArrayList<Double>> sentence_vectors, ArrayList<Double> the_vector)
{
    ArrayList<Double> scores=new ArrayList<Double>();
    ArrayList<Double> the_vector_normal=getNormalVec(the_vector);

    for(int sentence_index=0; sentence_index<sentence_vectors.size();
sentence_index++){
        ArrayList<Double>
sentence_normal=getNormalVec(sentence_vectors.get(sentence_index));
        double temp=0;
        for(int i=0;i<the_vector_normal.size();i++){

            temp=temp+the_vector_normal.get(i)*sentence_normal.get(i);

        }
    }
}

```



```

        scores.add(temp);
    }
    return scores;
}

```

viii. Sentence Similarity to the Average Sentence Similarity

After making the vectors for all the sentences, I make up an average vector based on their values that can show what the content approximately is. I then calculate the similarity of the all other sentences to this sentences.

```

static ArrayList<Double> make_average_vector(ArrayList<String> sorted_words,
ArrayList<Double> sorted_scores, ArrayList<String> inputlines){

```

```

    ArrayList<Double> average_vector=new ArrayList<Double>();
    for(int term=0;term<sorted_words.size();term++){
        average_vector.add(0.0);
    }
    ArrayList<ArrayList<Double>> sentence_vectors=new
ArrayList<ArrayList<Double>>();
    for(int
sentence_index=0;sentence_index<inputlines.size();sentence_index++){
        ArrayList<Double> sentence= new ArrayList<Double>();

        for(int term=0;term<sorted_words.size();term++){
            if(
(inputlines.get(sentence_index)).contains(sorted_words.get(term))){
                sentence.add(sorted_scores.get(term));
                double
current=average_vector.get(term)*(sentence_index);
                average_vector.set(term,
((current+sorted_scores.get(term))/(sentence_index+1)));
            }else{
                sentence.add(0.0);
            }
        }
        sentence_vectors.add(sentence);
    }
    return average_vector;
}

```

```

private static ArrayList<Double> calculate_similarity_score(
ArrayList<ArrayList<Double>> sentence_vectors, ArrayList<Double> the_vector)
{

```

```

    ArrayList<Double> scores=new ArrayList<Double>();
    ArrayList<Double> the_vector_normal=getNormalVec(the_vector);

    for(int
sentence_index=0;sentence_index<sentence_vectors.size();sentence_index++){
        ArrayList<Double>
sentence_normal=getNormalVec(sentence_vectors.get(sentence_index));
        double temp=0;
        for(int i=0;i<the_vector_normal.size();i++){

            temp=temp+the_vector_normal.get(i)*sentence_normal.get(i);

```

```
    }  
    scores.add(temp);  
}  
return scores;  
}
```

3) Sample Trace of Running Score Modules:

Query: wife

Input:

<h1>Sir Paul McCartney to marry for third time</h1>

<h2>Sir Paul McCartney is to marry his girlfriend of four years.</h2>

The former Beatle, who is 68, has become engaged to 51-year-old New York businesswoman Nancy Shevell, who will become his third wife.

Sir Paul's publicist Stuart Bell did not give any further details but said: "We're all thrilled for him." The date of the wedding is not yet known.

Sir Paul's first wife, Linda, died of cancer in 1998. He married Heather Mills in 2002 and divorced in 2008.

Ms Shevell was married for more than 20 years to American lawyer and political candidate Bruce Blakeman.

She is a board member of New York's transportation authority and a vice-president of her family business.

Output:

sir paul mccartney is to marry his girlfriend of four years. 4.093987097573501

sir paul mccartney to marry for third time 5.179668399358844

C:\Elnaz\Projects\workspace\Summerization\input1.txt

C:\Elnaz\Projects\workspace\Summerization\input2.txt

C:\Elnaz\Projects\workspace\Summerization\input3.txt

C:\Elnaz\Projects\workspace\Summerization\input4.txt

C:\Elnaz\Projects\workspace\Summerization\input5.txt

C:\Elnaz\Projects\workspace\Summerization\input6.txt

C:\Elnaz\Projects\workspace\Summerization\input7.txt

C:\Elnaz\Projects\workspace\Summerization\input8.txt

C:\Elnaz\Projects\workspace\Summerization\input9.txt

C:\Elnaz\Projects\workspace\Summerization\input10.txt

Query words: wife C:\Elnaz\Projects\workspace\Summerization\input10.txt

84

{1998=1, 20=1, 2002=1, 2008=1, 51-year-old=1, 68=1, a=2, all=1, american=1, and=3, any=1, authority=1, beatle=1, become=2, bell=1, blakeman=1, board=1, bruce=1, business=1, businesswoman=1, but=1, cancer=1, candidate=1, date=1, details=1, did=1, died=1, divorced=1, engaged=1, family=1, first=1, for=3, former=1, four=1, further=1, girlfriend=1, give=1, has=1, he=1, heather=1, her=1, him=1, his=2, in=3, is=4, known=1, lawyer=1, linda=1, married=2, marry=2, mccartney=2, member=1, mills=1, more=1, ms=1, nancy=1, new=2, not=2, of=5, paul=4, political=1, publicist=1, said=1, she=1, shevell=2, sir=4, stuart=1, than=1, the=3, third=2, thrilled=1, time=1, to=4, transportation=1, vice-president=1, was=1, we're=1, wedding=1, who=2, wife=2, will=1, years=2, yet=1, york=2}

370

{-=5, 0=3, 000=1, 0001=1, 0005=1, 05%=2, 1=1, 100=1, 2=2, 4=1, 7mm=1, 90=1, a=25, able=1, about=1, above=1, across=1, act=1, aeroplanes=1, ago=1, akin=1, almost=1, also=1, amount=1, an=4, and=5, answer=1, apart=1, appears=1, are=4, around=1, arrival=1, astronomers=1, at=5, atomic=1, australia=2, bank=3, bbc=2, be=5, beam=2, beams=3, because=2, been=1, behaved=1, being=2, between=1, big=1, black=2, blew=1, both=1, breaking=2, but=5, by=7, byrd=1, c=1, called=1, can=2, causing=1, city=1, clock=4, clocks=1, close=1, co-researcher=1, compact=1, compacted=1, companion=1, consists=1, constellation=1, cores=2, curvature=1, curved=4, day=1, dead=2, deeper=1, delay=2, delay'=1, dense=1, describes=1, description=1, detect=1, detected=1, dick=1, direction=1, directly=2, discovered=1, distance=2, distant=1, disturbed=1, double=5, dr=1, duncan=1, each=6, earth=5, eclipse=1, edge-on=1, effect=1, effects=1, einstein=6, emit=1, emitted=1, emitting=1, energy=1, environment=1, even=1, every=2, example=1, excellent=1, exhausted=1, existence=1, explain=1, explained=1, extra=1, extreme=1, extremely=1, facility=1, fall=1, far=2, field=1, final=1, find=1, flown=1, for=3, found=2, from=6, fuel=1, further=1, future=1, general=3, generated=1, giant=1, give=1, going=1, good=1, gr=2, gradually=1, gravitational=4, gravity=3, green=1, hapiro=1, happens=1, has=2, have=4, he=1, holes=2, hopes=1, hour=1, hours=1, how=1, ideas=2, imprint=1, in=14, indirect=1, international=1, into=1, is=5, it=8, jodrell=2, km=1, know=1, kramer=4, laboratory=1, large-scale=1, larger=1, latest=1, leader=1, least=1, leave=1, light=1, light-years=1, lighthouse-like=1, like=2, likely=1, lorimer=1, lose=1, lot=1, lovell=1, make=2, manchester=2, mass=2, match=1, mean=1, measure=2, measurements=1, mechanics=1, member=1, merging=1, michael=1, million=1, millionths=1,

minus=1, monitoring=1, more=3, morgantown=1, national=1, near=1, near-perfect=1, nearby=1, need=1, network=1, neutron=6, news=2, no=1, not=1, noted=1, nuclear=1, objects=1, observations=1, observatory=1, observed=4, of=27, on=3, one=3, or=1, orbit=2, other=7, our=2, out=1, over=1, pair=1, parkes=1, particles=1, pass=1, passed=1, passes=1, past=1, per=1, phenomenon=1, physical=1, plus=1, point=2, precisely=2, precision=1, predicted=1, predictions=1, predicts=1, pretty=1, probably=1, probing=1, problems=1, professor=4, pronounced=1, proof=1, properties=2, provide=1, pulsar=14, pulse=1, pulses=2, puppis=1, put=1, quantum=1, radio=5, range=1, rapidly=1, ratio=1, realm=1, received=1, recognise=1, referred=1, relativity=3, remarkable=1, remnant=1, reported=1, research=1, researchers=1, result=1, results=1, revision=1, ripples=1, robert=1, rotating=1, rule=1, run=2, said=2, saw=1, science=1, scientific=1, scientists=1, scrutiny=1, second=1, see=2, seek=1, seen=1, should=3, shrink=1, signal=4, sits=1, sitting=1, slower=1, so=2, some=1, space=3, space-time=6, speed=1, speeds=1, spiral=1, spread=1, stable=1, star=3, stars=5, stellar=2, stood=1, studied=1, studying=1, sub-atomic=1, such=1, sun=1, supermassive=1, sweep=1, system=6, team=8, telescope=5, tell=1, tell-tale=1, test=2, than=2, that=9, the=67, their=5, themselves=1, theory=4, these=2, they=6, this=3, three=2, through=3, thus=1, ticks=2, time=1, time-keeping=1, timepiece=1, to=22, told=2, towards=1, travel=1, two=3, uk=2, ultra-precise=1, underlying=1, unify=1, unifying=1, universe=2, university=2, up=1, us=3, use=1, used=1, using=1, values=1, very=1, virginia=2, was=1, waves=6, way=2, we=6, we've=1, weigh=1, well=2, were=2, west=2, what=1, when=1, whenever=1, whose=1, wide=1, will=2, with=2, within=1, would=1, years=2, yet=1}

309

{ \$2bn-\$3bn=1, \$30bn=1, -=3, 1=1, 2bn-£1=1, 300=1, 500=1, 700=1, 8bn=1, a=7, action=1, activities=1, after=1, agreed=1, aid=1, alain=1, all=1, alone=1, also=2, ambassador=1, amid=1, an=1, and=11, announced=2, any=1, arab=1, are=1, around=1, arrest=1, arrived=1, assets=2, assist=1, assistance=1, at=2, available=1, based=1, basic=2, bbc=1, be=8, because=1, become=1, bedraggled=1, been=1, before=1, began=1, behaviour=1, being=2, believed=1, benghazi=4, besieged=1, between=1, beyond=1, britain=1, british=1, but=1, by=3, call=1, can=1, caught=1, channelled=1, chartered=1, chief=1, city=1, civilians=1, clinton=1, col=3, coming=1, conflict=2, contact=3, contrary=1, could=1, council=2, countries=1, country=2, court=1, create=1, creates=1, creation=1, criminal=1, cut-off=1, date=1, declared=1, despite=1, details=1, did=2, diplomats=2, discuss=1, discussing=1, docked=1, don't=1, during=1, east=2, eastern=1, educating=1, effectively=1, efforts=1, enforcing=1, evacuated=1, exactly=1, expected=1, expelled=2, expelling=1, expenses=1, explain=1, explore=1, exporting=1, families=1, february=1, finally=1, financial=2, five=1, followed=1, food=1, for=6, forces=1, foreign=5, franco=1, frattini=1, free=1, french=1, from=3, frozen=1, fund=5, further=1, gaddafi=4, going=1, government=3, grata=1, group=5, had=3, hague=3, has=3, have=4, he=5, help=2, hillary=1, hit=1, hold=2, how=1, i=1, importing=1, in=13, increase=1, individuals=1, insisted=1, interests=1, international=3, iom=1, is=2, it=4, italian=1, italy=1, john=1, judged=1, juppe=1, keep=2, killed=2, last=1, later=1, libya=5, libyan=5, luis=1, made=1, make=1, managed=1, mandate=1, meanwhile=3, mechanism=1, medicine=1, meet=1, meeting=1, migration=1, military=2, minister=2, misrata=1, moment=1, money=1, month=1, months=1, more=2, moreno-ocampo=1, most=1, mostly=1, move=1, moving=1, mr=1, muammar=1, much=2, name=1, national=1, nations=1, nato=2, need=1, needs=1, new=1, non=1, not=3, now=1, of=16, officials=1, oil=1, on=5, operational=1, or=1, ordered=1, organization=1, other=2, out=1, owned=1, panic=1, passengers=1, people=6, permit=1, persona=1, port=1, portion=1, pressure=1, prevented=1, products=1, prosecutor=1, protect=1, public=1, re-

dock=1, rebel=6, rebels=1, red=1, refined=1, report=1, request=2, revenue=1, rome=2, said=9, sail=1, salaries=1, saying=1, says=3, scenes=1, secretary=3, security=1, seeking=1, separated=1, services=2, setting=1, she=1, shellfire=1, ship=3, should=2, so=1, some=1, speaking=1, spent=1, star=1, state=2, states=1, still=1, stranded=1, sudworth=1, supplies=1, take=1, talks=1, tap=1, tax=1, temporary=1, than=1, that=4, the=41, their=1, them=2, therefore=1, these=1, they=6, this=1, those=2, thursday=1, tired=1, to=23, top=1, transitional=1, transparently=1, trying=1, twice=1, two=1, uk=1, un=2, unacceptable=1, united=1, urgent=2, us=2, wanted=1, warrants=1, was=4, ways=1, we=1, weapons=1, wednesday=1, weekend=1, weeks=1, were=5, west=1, while=1, why=1, will=1, william=2, with=1, within=1, workers=1, would=4, wounded=1, £1=1}

181

{200=1, 2006=1, a=14, about=3, according=1, added=1, after=2, age=1, all=2, almost=1, and=6, antenatal=1, at=2, auto-immunity=1, babies=2, baby=1, baseline=1, bbc=1, be=5, been=1, before=1, benefits=1, beyond=1, birth=2, blood=3, but=3, by=1, call=1, can=1, care=1, charles=1, childbearing=1, college=1, comprehensive=1, congress=1, could=2, countries=2, czech=1, detected=1, detection=1, develop=2, direction=1, discussed=1, disease=3, dr=2, earlier=1, early=3, eliska=1, endocrinology=1, european=2, every=1, evidence=3, family=1, fine=1, followed=1, for=9, from=1, full-blown=1, function=1, future=1, gets=1, getting=1, gives=1, giving=1, gland=1, go=1, good=1, gp=1, grounds=1, gynaecologist=1, had=2, half=1, has=2, have=7, health=2, her=3, hidden=1, high-risk=1, history=1, if=1, immediately=1, implications=2, improve=1, in=10, including=1, is=2, it=1, jacobs=1, lack=1, larger=1, led=1, major=2, many=1, marker=1, may=1, merits=1, midwife=1, midwives=2, monitor=1, more=2, most=1, mother=2, mothers=1, mothers-to-be=1, needed=3, no=2, of=22, on=4, only=1, or=2, outcomes=1, past=1, pick=1, positive=1, possible=1, potlukova=2, prague=1, pregnancy=5, pregnant=4, problems=5, programme=1, recommend=1, rejected=1, repeated=1, research=1, researchers=1, right=1, risk=1, royal=1, said=3, say=2, scale=1, screened=1, screening=6, she=5, should=3, signs=1, step=1, study=1, sue=1, suffered=1, suggests=1, sure=1, symptoms=2, teacher=1, tens=1, test=1, tested=2, that=3, the=16, these=2, they=3, thinking=1, third=2, this=1, thousands=1, through=1, throughout=1, thyroid=11, to=6, told=2, two=1, uk=3, universal=2, university=1, up=1, us=2, visit=1, was=3, we=1, went=1, which=1, who=3, will=2, within=1, woman=1, women=8, works=1, would=1, years=1, young=1}

143

{1=1, 11=1, 15=1, 1998=1, 2009=1, 30=1, 8=1, 8-10=1, a=9, abbey=1, after=2, aged=1, alberta=1, an=2, and=8, announcing=1, are=1, at=2, be=3, begin=1, believed=1, brother=1, by=1, california=1, cambridge=1, canada=6, canadian=1, capacity=1, capital=1, celebrations=1, charles=1, clarence=1, close=1, commonwealth=1, cornwall=1, country=2, couple=3, date=1, day=2, duchess=4, duke=2, earlier=1, edward=1, extend=1, family=1, father=1, first=5, following=1, for=1, friday=1, from=2, full=1, harper=1, harry=1, has=1, helicopter=1, his=3, honeymoon=1, house=1, in=6, including=1, is=1, island=1, it=3, itinerary=1, its=1, job=1, joint=1, july=4, june=1, last=1, later=2, location=1, married=1, marrying=1, new=1, newly-weds=1, northwest=1, of=3, official=3, on=3, opted=1, ottawa=2, overseas=2, page=1, part=1, pay=1, pilot=1, pm=1, postpone=1, previously=1, prince=7, private=1, published=1, quebec=1, queen=1, raf=1, relationship=1, rescue=1, returned=1, royal=2, said=3, search=1, secret=1, speaking=1, spent=2, spokesman=1, states=1, stephen=1, summer=1, take=1, territories=1, testament=1, the=23, their=5, there=1, they=1, this=4, three-day=1, to=12, tour=2, touring=2, trip=3, twitter=1, uk=1, until=1, us=3, vancouver=1, very=1, visit=6, visited=3,

wales=1, was=1, week=1, weekend=1, westminster=1, while=1, wife=1, will=5,
william=2, with=3, would=1, year=1}

121

{\$1=1, \$3=1, \$36=1, \$49=1, \$900m=1, 12%=1, 15%=1, 2008=1, 2010=2, 25%=1,
2bn=2, 5bn=1, 7bn=1, 95bn=1, a=2, accompanied=1, also=2, america=1, and=4,
as=1, automotive=1, bankruptcy=2, beat=1, better=1, boosted=1, by=4,
capacity=1, car=1, carmaker=1, china=1, coming=1, consecutive=1, costs=1,
cuts=1, delphi=1, division=1, during=1, earnings=1, eventually=1,
excluding=1, expectations=1, expected=1, factory=1, fifth=1, figure=1,
financial=1, first=2, fixed=1, for=1, former=1, from=2, full=2, full-year=1,
funding=1, general=2, given=1, gm=4, government=1, has=1, helped=2, higher=2,
improvement=1, in=7, including=1, into=1, it=4, items=1, its=4, last=2,
lower=1, major=1, market=1, more=1, motor=2, north=1, of=6, one-off=1,
other=1, out=1, parts=1, pricing=1, proceeds=1, profit=6, profitable=1,
protection=1, quarter=2, recovery=1, reported=1, represent=1,
restructuring=1, results=3, revenues=1, rise=1, rose=1, said=1, sale=3,
sales=2, set=1, show=1, since=1, solid=1, staff=1, stake=1, strong=1,
surges=1, than=2, that=1, the=7, three=1, times=1, to=2, up=1, us=3, was=3,
went=1, were=3, which=1, with=1, worldwide=1, year=3, £1=1}

104

{10=1, 31=1, 4=1, 71=1, a=5, actor=2, ahead=1, al=2, also=1, amazing=1, an=2,
and=5, announced=1, at=3, august=1, award=3, begins=1, behind=1, broadway=1,
called=1, camera=1, cinema=1, contemporary=2, describes=1, director=3,
documentary=1, earlier=1, enriches=1, ever=1, experience=1, explores=1,
feature=1, festival=5, film=5, film-maker=2, for=1, given=1, glory=1,
godfather=1, has=1, have=1, he=1, herod=1, his=3, in=2, is=3, japan=1,
joins=1, kitano=1, left=1, marco=1, mark=2, merchant=1, most=1, mueller=1,
new=1, nomination=1, of=7, on=3, organisers=1, original=2, oscar=1, oscar-
winning=1, pacino=5, personal=1, play=2, precious=1, premiere=1, previous=1,
prize=1, production=1, project=1, receive=3, recieved=1, role=2, runs=1,
salome=2, september=2, shakespeare=1, special=1, stallone=1, star=1, stint=1,
such=1, sylvester=1, takeshi=1, the=18, third=1, this=2, to=4, tony=1,
trilogy=1, until=1, us=1, venice=4, week=1, which=1, who=2, whose=1, wilde=2,
will=1, winners=1, world=1, year=1}

351

{&=1, -=4, 11=1, 2013=1, a=27, abbotabad=1, about=2, account=2,
accusations=1, acknowledge=1, act=2, action=1, additional=1,
administration=1, admitted=1, after=3, again=3, agency=1, aircraft=1, al=1,
al-qaeda=2, all=1, allegations=1, along=1, also=2, an=5, and=9, any=2,
arabia=1, are=5, area=1, armed=1, army=2, around=1, arrest=1, ashfaq=1, at=6,
attacks=1, barbara=1, bashir=1, bbc=2, be=1, been=5, bin=18, blue=1, body=3,
building=1, buried=1, but=2, by=4, call=1, calls=2, capture=1, carrier=1,
changed=1, chief=1, claimed=1, climbing=1, co-operation=1, comes=1,
comments=1, completion=1, compound=4, comprises=1, concerned=1, concerns=1,
construction=1, continuing=1, contradictions=1, contributed=1, cooperation=1,
correspondent=2, could=1, country=2, courier=1, critical=1, critics=1, day=1,
dead=2, decided=1, decision=1, denied=1, details=1, did=1, diplomatic=1,
dismissed=1, effective=1, elsewhere=1, emerged=1, explain=1, failing=1,
failure=2, families=1, feelings=1, financial=1, find=1, fire=1, fire-fight=1,
fired=1, firefighters=1, first=1, five=1, flags=1, floating=1, for=6,
forces=3, forty=1, from=3, gathered=1, gen=1, given=1, going=1, graphic=2,
ground=3, had=7, has=5, hasan=1, have=2, he=6, head=2, her=1, highly=1,
his=4, horrible=1, house=3, however=3, husband=1, if=1, images=1,
important=1, impression=1, in=17, incitement=1, including=1, individual=1,

information=1, institution=1, into=1, investigated=1, investigation=2, investigators=1, is=4, it=7, justice=1, kayani=1, killed=8, killing=1, knowledge=1, laden=18, laid=2, last=1, lasting=1, launch=1, lawful=1, leader=2, leading=1, legality=1, level=1, links=1, living=1, locate=2, long=1, made=4, make=2, man=1, manhattan=1, meanwhile=1, media=1, memorial=1, memories=1, met=1, midtown=1, military=6, military/intelligence=1, minute=1, mixed=1, monday=2, more=1, most=1, mr=5, museum=2, national=5, new=6, night=1, no=1, northern=1, not=7, now=1, obama=7, of=24, office=1, official=1, officials=1, on=7, one=4, only=1, operation=2, opinion=1, or=2, osama=1, others=1, our=2, outlets=1, over=1, painful=1, pakistan=8, pakistani=3, park=1, people=1, perceived=1, perpetrators=1, phone=1, photographs=1, photos=3, pictures=1, plett=1, politicising=1, pose=1, presence=1, president=4, prior=1, priority=1, proof=1, propaganda=1, providing=1, public=3, publish=2, qaeda=1, quoted=1, raid=5, raided=1, raids=1, raised=1, rawalpindi=2, recently=1, recriminations=1, red=1, referring=1, relatives=1, reopen=1, respected=1, reuters=1, reveal=1, review=2, revised=1, risk=1, room=1, said=9, salman=1, same=1, saudi=2, saying=2, says=4, scene=1, scheduled=1, sea=1, secret=1, security=2, see=1, seen=1, self-defence=1, senior=1, sensitive=1, september=1, service=2, several=1, she=1, shoaib=1, shoot-out=1, shortcomings=1, shot=2, signs=1, similar=1, site=2, some=2, somebody=1, sovereignty=2, special=2, spoke=1, spoken=1, spokesman=1, staff=1, start=1, states=1, station=1, stop=1, street=1, such=1, sunday=1, sure=2, syed=1, target=1, that=10, the=58, their=1, then=1, there=5, they=1, this=1, those=1, thousands=1, throughout=1, thursday=2, times=1, to=16, told=3, tool=1, top=1, towers=1, traced=1, transactions=1, tuesday=1, unarmed=2, unilateral=1, united=1, unstable=1, us=9, very=2, victims=1, violating=1, violation=1, violence=1, visit=2, volatile=1, want=1, warned=1, warrant=1, was=14, waterfalls=1, waving=1, we=2, well=1, were=3, what=1, when=2, where=1, whereabouts=1, which=1, while=1, white=3, who=3, whose=1, why=1, will=4, with=4, wives=1, would=4, wreath=2, years=1, york=4, zero=3}

374

{-=5, 13%=1, 150-sq-km=1, 170km=1, 1775=2, 19th=1, 20th=1, 2kg=1, 50=1, 6m=1, a=16, about=2, abundant=1, accidental=1, accomplished=1, accrue=1, across=1, after=1, against=1, agent=1, ahead=1, aircraft=1, albatrosses=2, alive=1, all=1, along=2, also=1, although=1, anas=1, and=20, animals=1, antarcticus=1, anthus=1, anything=1, appear=1, are=2, area=3, around=2, at=2, atlantic=1, attempted=1, attractive=1, australia=1, back=3, bait=5, baited=1, baiting=2, base=1, batches=1, bbc=1, be=8, become=1, been=4, before=1, being=2, believe=1, bell=1, beneath=1, benefits=1, biggest=1, birds=4, breeding=1, british=1, brown=1, but=7, by=9, calculates=1, campaign=4, can=4, cannot=1, captain=1, carefully=1, centred=1, centuries=2, chairs=1, chicks=2, cleared=1, clock=1, coast=2, collateral=1, colour=1, committee=1, compared=1, completely=1, concern=1, conservationists=2, consuming=1, contained=1, convenient=1, cook=2, cost=1, could=1, cover=1, covered=1, cross=1, damage=1, date=1, days=2, dead=1, deaths=1, designed=1, devastating=1, divide=1, do=1, doing=1, donated=1, down=1, dr=1, driven=1, duck=1, dundee=1, early=1, eat=1, efforts=1, encouraged=1, ensuring=1, eradication=3, even=1, every=1, everywhere=1, explained=1, explorer=1, famed=1, far=1, fat=1, few=1, first=5, flew=1, foot=1, for=4, form=2, forth=1, found=2, from=4, further=1, georgia=12, georgica=2, get=2, glaciers=2, gone=1, gps=1, great=1, grid=1, ground=2, ground-nesting=1, habitat=1, had=2, has=3, have=7, he=1, hectare=1, helicopters=3, hemmed=1, heritage=3, history=1, hoppers=1, how=1, hugely=1, ice=1, if=1, impact=1, important=1, impossible=1, in=18, includes=1, industrial=1, industries=1, inevitably=1, inhabit=1, inspections=1, into=3, introduced=2, is=6, island=11, it=7, james=1, just=1, kill=3, killing=1,

knows=1, laboratories=1, land=1, landmass=1, large=1, later=1, laying=1,
 leader=1, leapt=1, legacy=1, lines=1, littered=1, little=1, local=1, long=2,
 losses=1, mainland=1, majority=1, man=1, many=1, march=2, marked=1, martin=2,
 may=1, mere=1, mike=1, millions=1, minimise=1, more=1, most=2,
 neighbouring=1, nest=1, nesting=1, new=2, night=1, no=1, no-go=1, no-one=1,
 nonetheless=1, north=1, not=3, notably=1, nothing=2, number=1, numerous=1,
 of=20, off=1, offshore=1, on=12, one=3, only=2, ordinarily=1, other=3, out=2,
 over=1, overseas=1, part=1, particular=1, particularly=1, pellets=4,
 penguins=1, peninsula=1, perhaps=1, petrels=1, phase=2, pintail=1, pipit=1,
 planet=1, plants=1, pleased=1, populations=3, presence=1, present=1,
 prisons=1, prior=1, probably=1, processing=1, professor=2, project=4,
 protectorate=1, put=1, ran=1, rat=15, rat-infested=1, re-infest=1, really=1,
 refugees=1, remains=1, remove=1, removed=1, represented=1, require=1,
 restoration=1, results=1, richardson=1, rightly=1, riot=1, rodent=5,
 rodenticide=2, running=1, said=1, saw=1, say=3, scale=1, seabird=3,
 sealers=2, seals=1, seasons=1, see=1, seems=1, set=2, several=1, shape=1,
 ships=2, shore=1, sign=1, simply=1, single=1, sir=1, size=3, slung=1,
 small=3, so=1, some=1, something=1, songbird=1, south=13, southerly=1,
 species=1, spread=2, steering=1, sub-antarctic=1, subsequent=1, success=2,
 such=2, sure=1, surrounds=1, surveyed=1, taken=2, teeth=1, tend=1,
 territory=2, that=6, thatcher=1, the=69, their=5, them=3, then=1, there=3,
 these=5, they=2, this=2, thousands=1, times=1, to=16, told=1, tongues=1,
 tonnes=1, tony=1, tourists=1, toxic=1, tried=1, trust=3, try=1, turning=1,
 two=3, uk=2, university=1, up=1, us=1, used=2, visit=1, was=3, we=3, we're=1,
 week=1, went=2, were=5, whalers=2, what=2, when=1, where=1, which=3, who=1,
 wildlife=2, will=4, with=2, world=2, would=1, years=1, you=1, zealand=1,
 zone=3, £1=1, £5m=1}

376

{-=2, 000=2, 000m=1, 1=1, 10%=1, 10-year=1, 1950s=1, 2=1, 2006=1, 21=1, 30=1,
 400=1, 500m=1, a=10, able=2, about=2, adults=1, affect=1, afford=1, after=1,
 aid=1, also=2, amazingly=1, american=1, an=4, analysis=1, and=23, andrews=1,
 animal=2, animals=3, antarctic=1, appear=1, are=10, around=2, as=2, at=6,
 attached=1, australia=1, bbc=1, be=1, been=5, being=1, below=1, between=1,
 biggest=1, biological=1, biologist=1, biuw=2, boehme=2, both=1, bothered=1,
 boxes=1, breathe=1, bringing=1, british=1, but=4, by=2, can=2, can't=2,
 carried=1, case=1, cause=1, centred=1, changes=1, climate=2, collaborative=1,
 collected=1, collecting=1, computerised=1, computers=1, conditions=1,
 conservation=1, conveyor=2, could=1, critical=1, data=4, deal=1, decline=1,
 deeper=1, depth=2, detail=1, details=1, different=2, dive=2, diving=2,
 doubt=1, down=3, dramatically=1, during=1, dynamic=2, eat=1, effort=1,
 elephant=7, emerge=1, enabled=1, end=1, energy=1, environment=1, epoxy=1,
 equipped=1, even=1, exchange=1, explain=1, explained=1, explaining=1, fall=2,
 fared=1, fedak=1, feed=1, find=1, for=6, france=1, from=6, front=3,
 frontal=1, fronts=4, fur=1, gather=1, gathered=1, gatty=1, geophysical=1,
 georgia=3, giant=1, given=1, global=1, globe=1, glue=1, go=5, gone=1, good=1,
 great=1, group=2, habitat=1, habits=1, had=1, half=2, happens=1, has=7,
 have=9, hawaii=1, he=1, heads=2, heat=2, help=1, helped=1, helps=2, high=1,
 honolulu=1, hope=1, hour=1, how=2, hunting=1, improved=2, in=18,
 information=3, insights=1, instead=1, interesting=1, into=1, involved=1,
 is=12, islands=1, it=8, its=1, journeys=1, just=2, kerguelen=1, km=1,
 knowing=1, laboratory=1, large-scale=1, lars=1, less=1, levels=1, light=1,
 like=1, lions=1, loggers=1, logistically=1, macquarie=2, mammal=1, marine=2,
 martin=1, mass=1, may=2, meeting=1, metabolic=1, meteorology=1, might=1,
 mike=1, minutes=1, moment=1, most=1, moulting=1, move=1, movement=2, much=1,
 must=1, navigate=1, new=6, not=4, now=2, nutrients=1, ocean=13,
 oceanographers=1, oceanographic=1, oceanography=1, of=28, off=1, often=1,

on=8, one=1, only=1, opening=1, or=1, other=1, our=2, over=2, parallel=1, part=1, partnership=1, parts=1, perhaps=1, physical=1, places=1, plays=1, pop=1, population=1, populations=1, positions=2, preferring=1, prey=1, processes=1, produce=1, profiles=1, project=4, provided=1, prowess=1, rate=1, record=1, recruited=1, redistribute=1, reduce=1, region=2, relatively=1, remarkable=1, research=1, researcher=1, researchers=2, resolution=1, resources=1, rest=1, resting=1, results=1, role=1, run=1, running=1, said=4, salinity=2, sample=1, samplers=1, sampling=1, satellite=1, says=1, science=3, scientists=2, scotland=1, sea=2, seal=7, seals=4, seals?=1, seaos=1, season=1, see=1, sensors=1, shallower=1, sharks=1, ships=1, show=2, shutting=1, side=1, significantly=1, simply=1, since=1, sleep=1, smru=6, snapshot=2, so=3, some=1, something=1, south=3, southern=4, species=1, speculative=1, spend=1, squid=2, st=1, stable=1, still=1, stress=1, stuck=1, study=1, such=1, surface=1, survey=1, swim=2, system=3, tags=2, team=2, temperature=2, than=1, thanks=1, that=3, the=72, their=8, them=1, themselves=1, then=1, there=2, these=6, they=11, this=6, thousands=1, through=1, thrown=1, time=1, timeframe=1, to=26, told=1, too=1, towed=1, trace=1, tracked=1, transit=1, transmit=1, trends=1, tuna=1, two=2, understand=1, understanding=2, union=1, unit=1, university=1, unprecedented=1, up=2, upwelled=1, us=3, use=2, using=2, very=1, via=1, wake=1, wandering=1, want=1, warm=1, water=5, way=1, we=5, weather=1, website=1, week=1, well=1, what=4, when=2, where=5, which=1, why=2, will=1, window=1, with=4, within=2, world=1, would=1, year=3, years=2, you=2}

tf: {1998=0.007407407407407408, 20=0.007407407407407408, 2002=0.007407407407407408, 2008=0.007407407407407408, 51-year-old=0.007407407407407408, 68=0.007407407407407408, a=0.014814814814814815, all=0.007407407407407408, american=0.007407407407407408, and=0.022222222222222223, any=0.007407407407407408, authority=0.007407407407407408, beatle=0.007407407407407408, become=0.014814814814814815, bell=0.007407407407407408, blakeman=0.007407407407407408, board=0.007407407407407408, bruce=0.007407407407407408, business=0.007407407407407408, businesswoman=0.007407407407407408, but=0.007407407407407408, cancer=0.007407407407407408, candidate=0.007407407407407408, date=0.007407407407407408, details=0.007407407407407408, did=0.007407407407407408, died=0.007407407407407408, divorced=0.007407407407407408, engaged=0.007407407407407408, family=0.007407407407407408, first=0.007407407407407408, for=0.022222222222222223, former=0.007407407407407408, four=0.007407407407407408, further=0.007407407407407408, girlfriend=0.007407407407407408, give=0.007407407407407408, has=0.007407407407407408, he=0.007407407407407408, heather=0.007407407407407408, her=0.007407407407407408, him=0.007407407407407408, his=0.014814814814814815, in=0.022222222222222223, is=0.02962962962962963, known=0.007407407407407408, lawyer=0.007407407407407408, linda=0.007407407407407408, married=0.014814814814814815, marry=0.014814814814814815, mccartney=0.014814814814814815, member=0.007407407407407408, mills=0.007407407407407408, more=0.007407407407407408, ms=0.007407407407407408, nancy=0.007407407407407408, new=0.014814814814814815, not=0.014814814814814815, of=0.037037037037037035, paul=0.02962962962962963, political=0.007407407407407408, publicist=0.007407407407407408, said=0.007407407407407408, she=0.007407407407407408, shevell=0.014814814814814815, sir=0.02962962962962963, stuart=0.007407407407407408, than=0.007407407407407408, the=0.022222222222222223, third=0.014814814814814815, thrilled=0.007407407407407408,

time=0.007407407407407408, to=0.02962962962962963,
transportation=0.007407407407407408, vice-president=0.007407407407407408,
was=0.007407407407407408, we're=0.007407407407407408,
wedding=0.007407407407407408, who=0.014814814814814815,
wife=0.014814814814814815, will=0.007407407407407408,
years=0.014814814814814815, yet=0.007407407407407408,
york=0.014814814814814815}
if: {1998=1.2039728043259361, 20=1.6094379124341003, 2002=1.6094379124341003,
2008=1.2039728043259361, 51-year-old=1.6094379124341003,
68=1.6094379124341003, a=-0.0953101798043249, all=0.5108256237659907,
american=1.2039728043259361, and=-0.0953101798043249, any=0.9162907318741551,
authority=1.6094379124341003, beatle=1.6094379124341003,
become=0.9162907318741551, bell=1.2039728043259361,
blakeman=1.6094379124341003, board=1.6094379124341003,
bruce=1.6094379124341003, business=1.6094379124341003,
businesswoman=1.6094379124341003, but=0.22314355131420976,
cancer=1.6094379124341003, candidate=1.6094379124341003,
date=0.6931471805599453, details=0.6931471805599453, did=0.9162907318741551,
died=1.6094379124341003, divorced=1.6094379124341003,
engaged=1.6094379124341003, family=0.9162907318741551,
first=0.5108256237659907, for=-0.0953101798043249, former=1.2039728043259361,
four=1.6094379124341003, further=0.6931471805599453,
girlfriend=1.6094379124341003, give=1.2039728043259361, has=-
0.0953101798043249, he=0.22314355131420976, heather=1.6094379124341003,
her=0.9162907318741551, him=1.6094379124341003, his=0.6931471805599453, in=-
0.0953101798043249, is=0.0, known=1.6094379124341003,
lawyer=1.6094379124341003, linda=1.6094379124341003,
married=1.2039728043259361, marry=1.6094379124341003,
mccartney=1.6094379124341003, member=1.2039728043259361,
mills=1.6094379124341003, more=0.22314355131420976, ms=1.6094379124341003,
nancy=1.6094379124341003, new=0.22314355131420976, not=0.3566749439387324,
of=-0.0953101798043249, paul=1.6094379124341003,
political=1.6094379124341003, publicist=1.6094379124341003, said=0.0,
she=0.6931471805599453, shevell=1.6094379124341003, sir=1.2039728043259361,
stuart=1.6094379124341003, than=0.5108256237659907, the=-0.0953101798043249,
third=0.9162907318741551, thrilled=1.6094379124341003,
time=0.9162907318741551, to=-0.0953101798043249,
transportation=1.6094379124341003, vice-president=1.6094379124341003,
was=0.10536051565782635, we're=1.2039728043259361,
wedding=1.6094379124341003, who=0.5108256237659907, wife=1.2039728043259361,
will=0.0, years=0.3566749439387324, yet=1.2039728043259361,
york=1.2039728043259361}
tf-if: {1998=0.008918317069081009, 20=0.01192176231432667,
2002=0.01192176231432667, 2008=0.008918317069081009, 51-year-
old=0.01192176231432667, 68=0.01192176231432667, a=-0.0014120026637677763,
all=0.0037838935093777093, american=0.008918317069081009, and=-
0.0021180039956516642, any=0.006787338754623371,
authority=0.01192176231432667, beatle=0.01192176231432667,
become=0.013574677509246742, bell=0.008918317069081009,
blakeman=0.01192176231432667, board=0.01192176231432667,
bruce=0.01192176231432667, business=0.01192176231432667,
businesswoman=0.01192176231432667, but=0.0016529151949200723,
cancer=0.01192176231432667, candidate=0.01192176231432667,
date=0.005134423559703298, details=0.005134423559703298,
did=0.006787338754623371, died=0.01192176231432667,
divorced=0.01192176231432667, engaged=0.01192176231432667,
family=0.006787338754623371, first=0.0037838935093777093, for=-

0.0021180039956516642, former=0.008918317069081009, four=0.01192176231432667, further=0.005134423559703298, girlfriend=0.01192176231432667, give=0.008918317069081009, has=-7.060013318838881E-4, he=0.0016529151949200723, heather=0.01192176231432667, her=0.006787338754623371, him=0.01192176231432667, his=0.010268847119406597, in=-0.0021180039956516642, is=0.0, known=0.01192176231432667, lawyer=0.01192176231432667, linda=0.01192176231432667, married=0.017836634138162017, marry=0.02384352462865334, mccartney=0.02384352462865334, member=0.008918317069081009, mills=0.01192176231432667, more=0.0016529151949200723, ms=0.01192176231432667, nancy=0.01192176231432667, new=0.0033058303898401446, not=0.005284073243536776, of=-0.00353000665941944, paul=0.04768704925730668, political=0.01192176231432667, publicist=0.01192176231432667, said=0.0, she=0.005134423559703298, shevell=0.02384352462865334, sir=0.035673268276324034, stuart=0.01192176231432667, than=0.0037838935093777093, the=-0.0021180039956516642, third=0.013574677509246742, thrilled=0.01192176231432667, time=0.006787338754623371, to=-0.0028240053275355526, transportation=0.01192176231432667, vice-president=0.01192176231432667, was=7.804482641320471E-4, we're=0.008918317069081009, wedding=0.01192176231432667, who=0.0075677870187554185, wife=0.017836634138162017, will=0.0, years=0.005284073243536776, yet=0.008918317069081009, york=0.017836634138162017}

1: sentence tf idf score:[0.14505537106785252, 0.16267779713181135, 0.19567058316532188, 0.2274072224715072, 0.23317244461150907, 0.15578922733449302, 0.10085359211809608]
2: query sim:[0.0, 0.0, 0.33268663113089747, 0.0, 0.23899181784028387, 0.0, 0.0]
3: average sim:[0.9285562443270285, 0.9112723408463931, 0.31700396523085694, 0.7377511016892419, 0.7311073863947276, 0.18354491522520844, 0.132209946755066]

Number of sentences: 7

sorted_words: 84

* current_word_setk : 1

[of]

[of]

* current_word_setk : 1

[to]

[to]

* current_word_setk : 1

[and]

[and]

* current_word_setk : 1

[for]

[for]

* current_word_setk : 1

[in]

The following synsets contain 'in' or a possible base form of that text:

[in, inch, indium, In, atomic number 49, Indiana, Hoosier State, IN, inwards, inward]

* current_word_setk : 1

[the]

[the]

* current_word_setk : 1

[a]

The following synsets contain 'a' or a possible base form of that text:

[a, angstrom, angstrom unit, A, vitamin A, antiophthalmic factor, axerophthol, deoxyadenosine monophosphate, adenine, ampere, amp, letter a, type A, group A]

* current_word_setk : 1

[has]

The following synsets contain 'has' or a possible base form of that text:

[has, hour angle, HA, have, have got, hold, feature, experience, receive, get, undergo, own, possess, let, consume, ingest, take in, take, throw, make, give, induce, stimulate, cause, accept, suffer, sustain, give birth, deliver, bear, birth]

* current_word_setk : 1

[is]

The following synsets contain 'is' or a possible base form of that text:

[is, iodine, iodine, I, atomic number 53, one, 1, ace, single, unity, letter i, be, exist, equal, constitute, represent, make up, comprise, follow, embody, personify, live, cost]

* current_word_setk : 1

[said]

The following synsets contain 'said' or a possible base form of that text:

[said, state, say, tell, allege, aver, suppose, read, order, enjoin, pronounce, articulate, enounce, sound out, enunciate, aforesaid, aforementioned]

* current_word_setk : 1

[will]

The following synsets contain 'will' or a possible base form of that text:

[will, volition, testament, wish, bequeath, leave]

* current_word_setk : 1

[was]

The following synsets contain 'was' or a possible base form of that text:

[was, Washington, Evergreen State, WA, be, exist, equal, constitute, represent, make up, comprise, follow, embody, personify, live, cost]

* current_word_setk : 1

[but]

The following synsets contain 'but' or a possible base form of that text:

[but, merely, simply, just, only]

* current_word_setk : 1

[he]

The following synsets contain 'he' or a possible base form of that text:

[he, helium, He, atomic number 2]

* current_word_setk : 1

[more]

The following synsets contain 'more' or a possible base form of that text:

[more, More, Thomas More, Sir Thomas More, more than, to a greater extent]

* current_word_setk : 1

[new]

The following synsets contain 'new' or a possible base form of that text:

[new, recently, newly, freshly, fresh, unexampled, novel, raw, newfangled, young, New, Modern]

* current_word_setk : 1

[all]

The following synsets contain 'all' or a possible base form of that text:

[all, wholly, entirely, completely, totally, altogether, whole]

* current_word_setk : 1

[first]

The following synsets contain 'first' or a possible base form of that text:

[first, number one, number 1, beginning, commencement, outset, get-go, start, kickoff, starting time, showtime, offset, first base, first-class honours]

degree, first gear, low gear, low, firstly, foremost, first of all, first off, for the first time, 1st, inaugural, initiative, initiatory, maiden, world-class]

* current_word_setk : 1

[than]

[than]

* current_word_setk : 1

[date]

The following synsets contain 'date' or a possible base form of that text:

[date, day of the month, appointment, engagement, particular date, escort, date stamp, go steady, go out, see]

* current_word_setk : 1

[details]

The following synsets contain 'details' or a possible base form of that text:

[details, inside information, detail, item, point, particular, contingent]

* current_word_setk : 1

[further]

The following synsets contain 'further' or a possible base form of that text:

[further, foster, promote, advance, boost, encourage, farther, far]

* current_word_setk : 1

[she]

[she]

* current_word_setk : 1

[not]

The following synsets contain 'not' or a possible base form of that text:

[not, non]

* current_word_setk : 1

[years]

The following synsets contain 'years' or a possible base form of that text:

[years, old age, age, eld, geezerhood, long time, days, year, twelvemonth, yr, class]

* current_word_setk : 1

[any]

The following synsets contain 'any' or a possible base form of that text:

[any, whatever, whatsoever]

* current_word_setk : 1

[did]

The following synsets contain 'did' or a possible base form of that text:

[did, make, do, perform, execute, fare, make out, come, get along, cause, practice, practise, exercise, suffice, answer, serve, act, behave, manage, dress, arrange, set, coif, coiffe, coiffure]

* current_word_setk : 1

[family]

The following synsets contain 'family' or a possible base form of that text:

[family, household, house, home, menage, family unit, family line, folk, kinfolk, kinsfolk, sept, phratry, class, category, fellowship, kin, kinsperson, syndicate, crime syndicate, mob]

* current_word_setk : 1

[her]

[her]

* current_word_setk : 1

[time]

The following synsets contain 'time' or a possible base form of that text:

[time, clip, clock time, fourth dimension, meter, metre, prison term, sentence, clock]

* current_word_setk : 1

[who]

The following synsets contain 'who' or a possible base form of that text:

[who, World Health Organization, WHO]

* current_word_setk : 1

[1998]

[1998]

* current_word_setk : 1

[2008]

[2008]

* current_word_setk : 1

[american]

The following synsets contain 'american' or a possible base form of that text:

[amerish, American, American English, American language]

* current_word_setk : 1

[bell]

The following synsets contain 'bell' or a possible base form of that text:

[bell, doorbell, buzzer, toll, ship's bell, bell shape, campana, Bell, Melville Bell, Alexander Melville Bell, Vanessa Bell, Vanessa Stephen, Alexander Bell, Alexander Graham Bell, chime, gong]

* current_word_setk : 1

[former]

The following synsets contain 'former' or a possible base form of that text:

[former, erstwhile, old, onetime, quondam, sometime, late, previous, early, other]

* current_word_setk : 1

[give]

The following synsets contain 'give' or a possible base form of that text:

[give, spring, springiness, yield, afford, pay, hold, throw, have, make, gift, present, devote, render, return, generate, impart, leave, pass on, establish, sacrifice, pass, hand, reach, turn over, dedicate, consecrate, commit, apply, grant, move over, give way, ease up, feed, contribute, chip in, kick in, collapse, fall in, cave in, break, founder, open]

* current_word_setk : 1

[member]

The following synsets contain 'member' or a possible base form of that text:

[member, extremity, appendage, penis, phallus]

* current_word_setk : 1

[we're]

[we're]

* current_word_setk : 1

[yet]

The following synsets contain 'yet' or a possible base form of that text:

[yet, so far, thus far, up to now, hitherto, heretofore, as yet, til now, until now, even, still, eventually, sooner or later, in time, one of these days, however, nevertheless, withal, all the same, even so, nonetheless, notwithstanding]

* current_word_setk : 1

[his]

The following synsets contain 'his' or a possible base form of that text:

[his, hello, hullo, hi, howdy, how-do-you-do, Hawaii, Hawai'i, Aloha State, HI]

* current_word_setk : 1

[20]

The following synsets contain '20' or a possible base form of that text:

[20, twenty, XX, xx]

* current_word_setk : 1

[2002]

[2002]
 * current_word_setk : 1
 [51-year-old]
 [51-year-old]
 * current_word_setk : 1
 [68]
 The following synsets contain '68' or a possible base form of that text:
 [68, sixty-eight, lxxviii]
 * current_word_setk : 1
 [authority]
 The following synsets contain 'authority' or a possible base form of that text:
 [authority, authorization, authorisation, potency, dominance, say-so, assurance, self-assurance, confidence, self-confidence, sureness, agency, federal agency, government agency, bureau, office, sanction]
 * current_word_setk : 1
 [beatle]
 [beatle]
 * current_word_setk : 1
 [blakeman]
 [blakeman]
 * current_word_setk : 1
 [board]
 The following synsets contain 'board' or a possible base form of that text:
 [board, plank, display panel, display board, gameboard, table, control panel, instrument panel, control board, panel, circuit board, circuit card, card, plug-in, add-in, dining table, get on, room]
 * current_word_setk : 1
 [bruce]
 The following synsets contain 'bruce' or a possible base form of that text:
 [bruce, Bruce, David Bruce, Sir David Bruce, Robert the Bruce, Robert I]
 * current_word_setk : 1
 [business]
 The following synsets contain 'business' or a possible base form of that text:
 [business, concern, business concern, business organization, business organisation, commercial enterprise, business enterprise, business sector, occupation, job, line of work, line, stage business, byplay, clientele, patronage]
 * current_word_setk : 1
 [businesswoman]
 The following synsets contain 'businesswoman' or a possible base form of that text:
 [businesswoman]
 * current_word_setk : 1
 [cancer]
 The following synsets contain 'cancer' or a possible base form of that text:
 [cancer, malignant neoplastic disease, Cancer, Crab, Cancer the Crab, genus Cancer]
 * current_word_setk : 1
 [candidate]
 The following synsets contain 'candidate' or a possible base form of that text:
 [candidate, campaigner, nominee, prospect]
 * current_word_setk : 1
 [died]
 The following synsets contain 'died' or a possible base form of that text:

[died, die, decease, perish, go, exit, pass away, expire, pass, kick the bucket, cash in one's chips, buy the farm, conk, give-up the ghost, drop dead, pop off, choke, croak, snuff it, fail, go bad, give way, give out, conk out, break, break down, die out, pall, become flat]

* current_word_setk : 1

[divorced]

The following synsets contain 'divorced' or a possible base form of that text:

[divorced, disassociate, dissociate, divorce, disunite, disjoint, split up]

* current_word_setk : 1

[engaged]

The following synsets contain 'engaged' or a possible base form of that text:

[engaged, prosecute, engage, pursue, absorb, engross, occupy, hire, employ, betroth, affiance, plight, wage, enlist, lease, rent, charter, take, mesh, lock, operate, occupied, booked, set-aside, busy, in use, meshed, intermeshed, affianced, bespoken, betrothed, pledged]

* current_word_setk : 1

[four]

The following synsets contain 'four' or a possible base form of that text:

[four, 4, IV, tetrad, quatern, quaternion, quaternary, quaternity, quartet, quadruplet, foursome, Little Joe, iv]

* current_word_setk : 1

[girlfriend]

The following synsets contain 'girlfriend' or a possible base form of that text:

[girlfriend, girl, lady friend]

* current_word_setk : 1

[heather]

The following synsets contain 'heather' or a possible base form of that text:

[heather, ling, Scots heather, broom, Calluna vulgaris, heather mixture]

* current_word_setk : 1

[him]

[him]

* current_word_setk : 1

[known]

The following synsets contain 'known' or a possible base form of that text:

[known, know, cognize, cognise, experience, live, acknowledge, recognize, recognise, roll in the hay, love, make out, make love, sleep with, get laid, have sex, do it, be intimate, have intercourse, have it away, have it off, screw, fuck, jazz, eff, hump, lie with, bed, have a go at it, bang, get it on, bonk]

* current_word_setk : 1

[lawyer]

The following synsets contain 'lawyer' or a possible base form of that text:

[lawyer, attorney]

* current_word_setk : 1

[linda]

[linda]

* current_word_setk : 1

[mills]

The following synsets contain 'mills' or a possible base form of that text:

[mills, Mills, Robert Mills, factory, mill, manufacturing plant, manufactory, Mill, James Mill, John Mill, John Stuart Mill, grinder, milling machinery, grind, pulverization, pulverisation, mill about, mill around]

* current_word_setk : 1

[ms]

The following synsets contain 'ms' or a possible base form of that text:

[ms, multiple sclerosis, MS, disseminated sclerosis, disseminated multiple sclerosis, Mississippi, Magnolia State, Master of Science, SM, MSc, manuscript, Ms, Ms., meter, metre, m, molarity, molar concentration, M, thousand, one thousand, 1000, K, chiliad, G, grand, thou, yard, letter m]
 * current_word_setk : 1
 [nancy]
 The following synsets contain 'nancy' or a possible base form of that text:
 [nancy, Nancy]
 * current_word_setk : 1
 [political]
 The following synsets contain 'political' or a possible base form of that text:
 [political]
 * current_word_setk : 1
 [publicist]
 The following synsets contain 'publicist' or a possible base form of that text:
 [publicist, publicizer, publiciser]
 * current_word_setk : 1
 [stuart]
 The following synsets contain 'stuart' or a possible base form of that text:
 [stuart, Stuart, Gilbert Stuart, Gilbert Charles Stuart]
 * current_word_setk : 1
 [thrilled]
 The following synsets contain 'thrilled' or a possible base form of that text:
 [thrilled, thrill, tickle, vibrate, shudder, shiver, throb, exhilarate, inebriate, exalt, beatify]
 * current_word_setk : 1
 [transportation]
 The following synsets contain 'transportation' or a possible base form of that text:
 [transportation, transportation system, transit, transfer, transferral, conveyance, fare, Department of Transportation, Transportation, DoT, shipping, transport, exile, deportation, expatriation]
 * current_word_setk : 1
 [vice-president]
 [vice-president]
 * current_word_setk : 1
 [wedding]
 The following synsets contain 'wedding' or a possible base form of that text:
 [wedding, wedding ceremony, nuptials, hymeneals, marriage, marriage ceremony, wedding party, marry, get married, wed, conjoin, hook up with, get hitched with, espouse, tie, splice]
 * current_word_setk : 1
 [become]
 The following synsets contain 'become' or a possible base form of that text:
 [become, go, get, turn, suit]
 * current_word_setk : 1
 [third]
 The following synsets contain 'third' or a possible base form of that text:
 [third, one-third, tierce, third base, third gear, thirdly, 3rd, tertiary]
 * current_word_setk : 1
 [married]
 The following synsets contain 'married' or a possible base form of that text:
 [married, marry, get married, wed, conjoin, hook up with, get hitched with, espouse, tie, splice, marital, matrimonial]

```

* current_word_setk : 1
[wife]
The following synsets contain 'wife' or a possible base form of that text:
[wife, married woman]
* current_word_setk : 1
[york]
The following synsets contain 'york' or a possible base form of that text:
[york, York, House of York]
* current_word_setk : 1
[marry]
The following synsets contain 'marry' or a possible base form of that text:
[marry, get married, wed, conjoin, hook up with, get hitched with, espouse,
tie, splice]
* current_word_setk : 1
[mccartney]
The following synsets contain 'mccartney' or a possible base form of that
text:
[mccartney, McCartney, Paul McCartney, Sir James Paul McCartney]
* current_word_setk : 1
[shevell]
[shevell]
* current_word_setk : 1
[sir]
The following synsets contain 'sir' or a possible base form of that text:
[sir, Sir]
* current_word_setk : 1
[paul]
The following synsets contain 'paul' or a possible base form of that text:
[paul, Paul, Alice Paul, Saint Paul, St. Paul, Apostle Paul, Paul the
Apostle, Apostle of the Gentiles, Saul, Saul of Tarsus]
* * * Just to check : 84
own    in both 2    3
is      in both 1    2
is      in both 1    3
1       in both 1    4
is      in both 1    6
is      in both 2    3
1       in both 2    4
is      in both 2    6
1       in both 3    4
is      in both 3    6
new     in both 2    6
date    in both 3    5
she     in both 2    5
she     in both 2    6
she     in both 5    6
age     in both 1    2
years   in both 1    5
age     in both 2    5
come    in both 2    3
come    in both 2    5
did     in both 3    5
her     in both 3    4
her     in both 3    6
her     in both 4    6
hi      in both 0    1
hi      in both 0    2

```

```

hi      in both 0   3
his     in both 1   2
20      in both 4   5
business in both 2   6
iv      in both 1   3
iv      in both 1   4
iv      in both 3   4
marry   in both 0   1
marry   in both 0   3
marry   in both 1   3
third   in both 0   2
marry   in both 0   1
wed     in both 0   3
marry   in both 0   4
marry   in both 0   5
wed     in both 1   3
marry   in both 1   4
marry   in both 1   5
wed     in both 3   4
wed     in both 3   5
married in both 4   5
wife    in both 2   4
york    in both 2   6
marry   in both 0   1
wed     in both 0   3
wed     in both 1   3
mccartney in both 0   1
shevell in both 2   5
sir     in both 0   1
sir     in both 0   3
sir     in both 0   4
sir     in both 1   3
sir     in both 1   4
sir     in both 3   4
paul    in both 0   1
paul    in both 0   3
paul    in both 0   4
paul    in both 1   3
paul    in both 1   4
paul    in both 3   4
0        10    6    9    5    5    2    =37.0 4.625
10       0     6    11   8    5    4    =44.0 4.0
6        6     0    11   7    10   10   =50.0 2.0
9        11    11    0    12   8    8    =59.0 1.84375
5        8     7    12   0    6    8    =46.0 1.9166666666666667
5        5     10   8    6    0    6    =40.0 2.3529411764705883
2        4     10   8    8    6    0    =38.0 2.1111111111111111

```

&&*****

final sentence scores based on scores : [4.093987097573501,
5.179668399358844, 2.9639619876190597, 3.9388665347491165,
3.6957545275466193, -8.139217961395333, -11.7330205854518]
final rank of sentences: [1, 0, 3, 4, 2, 5, 6]

[sir paul mccartney is to marry his girlfriend of four years., sir paul
mccartney to marry for third time]

Query: actress

Input 2:

<h1>West Side Story playwright Arthur Laurents dies</h1>

<h2>Arthur Laurents, writer of such classic stage musicals as West Side Story and Gypsy, has died in New York aged 93.</h2>

The director and screenwriter died at his Manhattan home from complications of pneumonia, his agent said.

Born in Brooklyn, the attorney's son began in radio and wrote military training films during World War II.

His screen credits include the Alfred Hitchcock film Rope, Barbra Streisand romance The Way We Were and 1977 ballet drama The Turning Point.

Laurents won a Tony award in 1968 as author of the book for the musical Hallelujah, Baby!, and another, in 1984, for directing La Cage aux Folles.

He remains best known for writing the books for West Side Story and Gypsy, hit Broadway shows that were later turned into movies.

Featuring music by Leonard Bernstein and lyrics by Stephen Sondheim, the former retold the Romeo and Juliet story as a drama about rival New York street gangs.

The latter, based on the memoirs of stripper Gypsy Rose Lee, premiered in 1959 and was successfully revived four times on Broadway.

Laurents directed three of the revivals himself, most recently in 2008 with Patty LuPone in the leading role.

His other credits as a stage director include I Can Get It For You Wholesale, best remembered as the musical which introduced a 19-year-old Barbra Streisand to Broadway in 1962.

Earlier this year the Oscar-winning actress confirmed she plans to star in and possibly direct a new film version of Gypsy.

west side story playwright arthur laurents dies 15.831650063202948

arthur laurents, writer of such classic stage musicals as west side story and gypsy, has died in new york aged 93. 7.08480888217564

laurents won a tony award in 1968 as author of the book for the musical hallelujah, baby!, and another, in 1984, for directing la cage aux folles. -3.680656369981719

Trace:

[illegible]

[illegible]

&=1, --4, 11=1, 2013=1, a=27, abbottabad=1, about=2, account=2, accusations=1, acknowledge=1, act=2, action=1, additional=1, administration=1, admitted=1, after=3, again=3, agency=1, aircraft=1, al=1, al-qaeda=2, all=1, allegations=1, along=1, also=2, an=5, and=9, any=2, arabia=1, are=5, area=1, armed=1, army=2, around=1, arrest=1, ashfaq=1, at=6, attacks=1, barbara=1, bashir=1, bbc=2, be=1, been=5, bin=18, blue=1, body=3, building=1, buried=1, but=2, by=4, call=1, calls=2, capture=1, carrier=1, changed=1, chief=1, claimed=1, climbing=1, co-operation=1, comes=1, comments=1, completion=1, compound=4, comprises=1, concerned=1, concerns=1, construction=1, continuing=1, contradictions=1, contributed=1, cooperation=1, correspondent=2, could=1, country=2, courier=1, critical=1, critics=1, day=1, dead=2, decided=1, decision=1, denied=1, details=1, did=1, diplomatic=1, dismissed=1, effective=1, elsewhere=1, emerged=1, explain=1, failing=1, failure=2, families=1, feelings=1, financial=1, find=1, fire=1, fire-fight=1, fired=1, firefighters=1, first=1, five=1, flags=1, floating=1, for=6, forces=3, forty=1, from=3, gathered=1, gen=1, given=1, going=1, graphic=2, ground=3, had=7, has=5, hasan=1, have=2, he=6, head=2, her=1, highly=1, his=4, horrible=1, house=3, however=3, husband=1, if=1, images=1, important=1, impression=1, in=17, incitement=1, including=1, individual=1, information=1, institution=1, into=1, investigated=1, investigation=2, investigators=1, is=4, it=7, justice=1, kayani=1, killed=8, killing=1, knowledge=1, laden=18, laid=2, last=1, lasting=1, launch=1, lawful=1, leader=2, leading=1, legality=1, level=1, links=1, living=1, locate=2, long=1, made=4, make=2, man=1, manhattan=1, meanwhile=1, media=1, memorial=1, memories=1, met=1, midtown=1, military=6, military/intelligence=1, minute=1, mixed=1, monday=2, more=1, most=1, mr=5, museum=2, national=5, new=6, night=1, no=1, northern=1, not=7, now=1, obama=7, of=24, office=1, official=1, officials=1, on=7, one=4, only=1, operation=2, opinion=1, or=2, osama=1, others=1, our=2, outlets=1, over=1, painful=1, pakistan=8, pakistani=3, park=1, people=1, perceived=1, perpetrators=1, phone=1, photographs=1, photos=3, pictures=1, plett=1, politicising=1, pose=1, presence=1, president=4, prior=1, priority=1, proof=1, propaganda=1, providing=1, public=3, publish=2, qaeda=1, quoted=1, raid=5, raided=1, raids=1, raised=1, rawalpindi=2, recently=1, recriminations=1, red=1, referring=1, relatives=1, reopen=1, respected=1, reuters=1, reveal=1, review=2, revised=1, risk=1, room=1, said=9, salman=1, same=1, saudi=2, saying=2, says=4, scene=1, scheduled=1, sea=1, secret=1, security=2, see=1, seen=1, self-defence=1, senior=1, sensitive=1, september=1, service=2, several=1, she=1, shoaib=1, shoot-out=1, shortcomings=1, shot=2, signs=1, similar=1, site=2, some=2, somebody=1, sovereignty=2, special=2, spoke=1, spoken=1, spokesman=1, staff=1, start=1, states=1, station=1, stop=1, street=1, such=1, sunday=1, sure=2, syed=1, target=1, that=10, the=58, their=1, then=1, there=5, they=1, this=1, those=1, thousands=1, throughout=1

1977=0.0037174721189591076, 1984=0.0037174721189591076,
2008=0.0037174721189591076, 93=0.0037174721189591076, a=0.03345724907063197,
about=0.0037174721189591076, actress=0.0037174721189591076,
aged=0.0037174721189591076, agent=0.0037174721189591076,
alfred=0.0037174721189591076, and=0.03717472118959108,
another=0.0037174721189591076, arthur=0.007434944237918215,
as=0.0037174721189591076, at=0.0037174721189591076,
attorney=0.0037174721189591076, author=0.0037174721189591076,
aux=0.0037174721189591076, award=0.0037174721189591076,
baby!=0.0037174721189591076, ballet=0.0037174721189591076,
barbra=0.007434944237918215, based=0.0037174721189591076,
began=0.0037174721189591076, bernstein=0.0037174721189591076,
best=0.007434944237918215, book=0.007434944237918215,
born=0.0037174721189591076, broadway=0.011152416356877323,
brooklyn=0.0037174721189591076, by=0.007434944237918215,
cage=0.0037174721189591076, can=0.0037174721189591076,
classic=0.0037174721189591076, complications=0.0037174721189591076,
confirmed=0.0037174721189591076, credits=0.007434944237918215,
died=0.007434944237918215, dies=0.0037174721189591076,
direct=0.0037174721189591076, directed=0.0037174721189591076,
directing=0.0037174721189591076, director=0.007434944237918215,
drama=0.007434944237918215, during=0.0037174721189591076,
earlier=0.0037174721189591076, featuring=0.0037174721189591076,
film=0.007434944237918215, films=0.0037174721189591076,
folles=0.0037174721189591076, for=0.01858736059479554,
former=0.0037174721189591076, four=0.0037174721189591076,
from=0.0037174721189591076, gangs=0.0037174721189591076,
get=0.0037174721189591076, gypsy=0.01486988847583643,
hallelujah=0.0037174721189591076, has=0.0037174721189591076,
he=0.0037174721189591076, himself=0.0037174721189591076,
his=0.01486988847583643, hit=0.0037174721189591076,
hitchcock=0.0037174721189591076, home=0.0037174721189591076,
i=0.0037174721189591076, ii=0.0037174721189591076, in=0.03717472118959108,
include=0.007434944237918215, into=0.0037174721189591076,
introduced=0.0037174721189591076, it=0.0037174721189591076,
juliet=0.0037174721189591076, known=0.0037174721189591076,
la=0.0037174721189591076, later=0.0037174721189591076,
latter=0.0037174721189591076, laurents=0.01486988847583643,
leading=0.0037174721189591076, lee=0.0037174721189591076,
leonard=0.0037174721189591076, lupone=0.0037174721189591076,
lyrics=0.0037174721189591076, manhattan=0.0037174721189591076,
memoirs=0.0037174721189591076, military=0.0037174721189591076,
most=0.0037174721189591076, movies=0.0037174721189591076,
music=0.0037174721189591076, musical=0.007434944237918215,
musicals=0.0037174721189591076, new=0.011152416356877323,
of=0.022304832713754646, on=0.007434944237918215, oscar-
winning=0.0037174721189591076, other=0.0037174721189591076,
patty=0.0037174721189591076, plans=0.0037174721189591076,
playwright=0.0037174721189591076, pneumonia=0.0037174721189591076,
point=0.0037174721189591076, possibly=0.0037174721189591076,
premiered=0.0037174721189591076, radio=0.0037174721189591076,
recently=0.0037174721189591076, remains=0.0037174721189591076,
remembered=0.0037174721189591076, retold=0.0037174721189591076,
revivals=0.0037174721189591076, revived=0.0037174721189591076,
rival=0.0037174721189591076, role=0.0037174721189591076,
romance=0.0037174721189591076, romeo=0.0037174721189591076,
rope=0.0037174721189591076, rose=0.0037174721189591076,

said=0.0037174721189591076, screen=0.0037174721189591076,
screenwriter=0.0037174721189591076, she=0.0037174721189591076,
shows=0.0037174721189591076, side=0.011152416356877323,
son=0.0037174721189591076, sonenheim=0.0037174721189591076,
stage=0.007434944237918215, star=0.0037174721189591076,
stephen=0.0037174721189591076, story=0.01486988847583643,
street=0.0037174721189591076, streisand=0.007434944237918215,
stripper=0.0037174721189591076, successfully=0.0037174721189591076,
such=0.0037174721189591076, that=0.0037174721189591076,
the=0.05947955390334572, this=0.0037174721189591076,
three=0.0037174721189591076, times=0.0037174721189591076,
to=0.007434944237918215, tony=0.0037174721189591076,
training=0.0037174721189591076, turned=0.0037174721189591076,
turning=0.0037174721189591076, version=0.0037174721189591076,
war=0.0037174721189591076, was=0.0037174721189591076,
way=0.0037174721189591076, we=0.0037174721189591076,
were=0.007434944237918215, west=0.011152416356877323,
which=0.0037174721189591076, wholesale=0.0037174721189591076,
with=0.0037174721189591076, won=0.0037174721189591076,
world=0.0037174721189591076, writer=0.0037174721189591076,
writing=0.0037174721189591076, wrote=0.0037174721189591076,
year=0.0037174721189591076, york=0.007434944237918215,
you=0.0037174721189591076}
if: {19-year-old=1.6094379124341003, 1959=1.6094379124341003,
1962=1.6094379124341003, 1968=1.6094379124341003, 1977=1.6094379124341003,
1984=1.6094379124341003, 2008=1.2039728043259361, 93=1.6094379124341003, a=-
0.0953101798043249, about=0.3566749439387324, actress=1.6094379124341003,
aged=1.2039728043259361, agent=1.2039728043259361, alfred=1.6094379124341003,
and=-0.0953101798043249, another=1.6094379124341003,
arthur=1.6094379124341003, as=0.9162907318741551, at=0.0,
attorney=1.6094379124341003, author=1.6094379124341003,
aux=1.6094379124341003, award=1.2039728043259361, baby!=1.6094379124341003,
ballet=1.6094379124341003, barbra=1.6094379124341003,
based=1.2039728043259361, began=1.2039728043259361,
bernstein=1.6094379124341003, best=1.6094379124341003,
book=1.6094379124341003, born=1.6094379124341003,
broadway=1.2039728043259361, brooklyn=1.6094379124341003, by=0.0,
cage=1.6094379124341003, can=0.3566749439387324, classic=1.6094379124341003,
complications=1.6094379124341003, confirmed=1.6094379124341003,
credits=1.6094379124341003, died=1.6094379124341003, dies=1.6094379124341003,
direct=1.6094379124341003, directed=1.6094379124341003,
directing=1.6094379124341003, director=1.2039728043259361,
drama=1.6094379124341003, during=0.6931471805599453,
earlier=0.6931471805599453, featuring=1.6094379124341003,
film=1.2039728043259361, films=1.6094379124341003, folles=1.6094379124341003,
for=-0.0953101798043249, former=1.2039728043259361, four=1.6094379124341003,
from=0.0, gangs=1.6094379124341003, get=1.2039728043259361,
gypsy=1.6094379124341003, hallelujah=1.6094379124341003, has=-
0.0953101798043249, he=0.22314355131420976, himself=1.6094379124341003,
his=0.6931471805599453, hit=1.2039728043259361, hitchcock=1.6094379124341003,
home=1.6094379124341003, i=1.2039728043259361, ii=1.6094379124341003, in=-
0.0953101798043249, include=1.6094379124341003, into=0.3566749439387324,
introduced=1.2039728043259361, it=0.0, juliet=1.6094379124341003,
known=1.6094379124341003, la=1.6094379124341003, later=0.6931471805599453,
latter=1.6094379124341003, laurents=1.6094379124341003,
leading=1.2039728043259361, lee=1.6094379124341003,
leonard=1.6094379124341003, lupone=1.6094379124341003,

lyrics=1.6094379124341003, manhattan=1.2039728043259361, memoirs=1.6094379124341003, military=0.9162907318741551, most=0.22314355131420976, movies=1.6094379124341003, music=1.6094379124341003, musical=1.6094379124341003, musicals=1.6094379124341003, new=0.22314355131420976, of=-0.0953101798043249, on=0.0, oscar-winning=1.2039728043259361, other=0.3566749439387324, patty=1.6094379124341003, plans=1.6094379124341003, playwright=1.6094379124341003, pneumonia=1.6094379124341003, point=1.2039728043259361, possibly=1.6094379124341003, premiered=1.6094379124341003, radio=1.2039728043259361, recently=1.2039728043259361, remains=1.2039728043259361, remembered=1.6094379124341003, retold=1.6094379124341003, revivals=1.6094379124341003, revived=1.6094379124341003, rival=1.6094379124341003, role=0.9162907318741551, romance=1.6094379124341003, romeo=1.6094379124341003, rope=1.6094379124341003, rose=1.2039728043259361, said=0.0, screen=1.6094379124341003, screenwriter=1.6094379124341003, she=0.6931471805599453, shows=1.6094379124341003, side=1.2039728043259361, son=1.6094379124341003, sondeim=1.6094379124341003, stage=1.6094379124341003, star=0.6931471805599453, stephen=1.2039728043259361, story=1.6094379124341003, street=1.2039728043259361, streisand=1.6094379124341003, stripper=1.6094379124341003, successfully=1.6094379124341003, such=0.3566749439387324, that=0.10536051565782635, the=-0.0953101798043249, this=0.0, three=0.9162907318741551, times=0.6931471805599453, to=-0.0953101798043249, tony=0.9162907318741551, training=1.6094379124341003, turned=1.6094379124341003, turning=1.2039728043259361, version=1.6094379124341003, war=1.6094379124341003, was=0.10536051565782635, way=0.9162907318741551, we=0.22314355131420976, were=0.3566749439387324, west=0.9162907318741551, which=0.22314355131420976, wholesale=1.6094379124341003, with=0.10536051565782635, won=1.6094379124341003, world=0.6931471805599453, writer=1.6094379124341003, writing=1.6094379124341003, wrote=1.6094379124341003, year=0.5108256237659907, york=1.2039728043259361, you=0.9162907318741551} tf-if: {19-year-old=0.005983040566669517, 1959=0.005983040566669517, 1962=0.005983040566669517, 1968=0.005983040566669517, 1977=0.005983040566669517, 1984=0.005983040566669517, 2008=0.004475735332066677, 93=0.005983040566669517, a=-0.003188816424680015, about=0.0013259291596235405, actress=0.005983040566669517, aged=0.004475735332066677, agent=0.004475735332066677, alfred=0.005983040566669517, and=-0.0035431293607555724, another=0.005983040566669517, arthur=0.011966081133339034, as=0.003406285248602807, at=0.0, attorney=0.005983040566669517, author=0.005983040566669517, aux=0.005983040566669517, award=0.004475735332066677, baby!=0.005983040566669517, ballet=0.005983040566669517, barbra=0.011966081133339034, based=0.004475735332066677, began=0.004475735332066677, bernstein=0.005983040566669517, best=0.011966081133339034, book=0.011966081133339034, born=0.005983040566669517, broadway=0.01342720599620003, brooklyn=0.005983040566669517, by=0.0, cage=0.005983040566669517, can=0.0013259291596235405, classic=0.005983040566669517, complications=0.005983040566669517, confirmed=0.005983040566669517, credits=0.011966081133339034, died=0.011966081133339034, dies=0.005983040566669517, direct=0.005983040566669517, directed=0.005983040566669517, directing=0.005983040566669517, director=0.008951470664133354, drama=0.011966081133339034, during=0.002576755318066711,

earlier=0.002576755318066711, featuring=0.005983040566669517,
film=0.008951470664133354, films=0.005983040566669517,
folles=0.005983040566669517, for=-0.0017715646803777862,
former=0.004475735332066677, four=0.005983040566669517, from=0.0,
gangs=0.005983040566669517, get=0.004475735332066677,
gypsy=0.02393216226667807, hallelujah=0.005983040566669517, has=-
3.543129360755572E-4, he=8.295299305360957E-4, himself=0.005983040566669517,
his=0.010307021272266844, hit=0.004475735332066677,
hitchcock=0.005983040566669517, home=0.005983040566669517,
i=0.004475735332066677, ii=0.005983040566669517, in=-0.0035431293607555724,
include=0.011966081133339034, into=0.0013259291596235405,
introduced=0.004475735332066677, it=0.0, juliet=0.005983040566669517,
known=0.005983040566669517, la=0.005983040566669517,
later=0.002576755318066711, latter=0.005983040566669517,
laurents=0.02393216226667807, leading=0.004475735332066677,
lee=0.005983040566669517, leonard=0.005983040566669517,
lupone=0.005983040566669517, lyrics=0.005983040566669517,
manhattan=0.004475735332066677, memoirs=0.005983040566669517,
military=0.003406285248602807, most=8.295299305360957E-4,
movies=0.005983040566669517, music=0.005983040566669517,
musical=0.011966081133339034, musicals=0.005983040566669517,
new=0.0024885897916082874, of=-0.002125877616453343, on=0.0, oscar-
winning=0.004475735332066677, other=0.0013259291596235405,
patty=0.005983040566669517, plans=0.005983040566669517,
playwright=0.005983040566669517, pneumonia=0.005983040566669517,
point=0.004475735332066677, possibly=0.005983040566669517,
premiered=0.005983040566669517, radio=0.004475735332066677,
recently=0.004475735332066677, remains=0.004475735332066677,
remembered=0.005983040566669517, retold=0.005983040566669517,
revivals=0.005983040566669517, revived=0.005983040566669517,
rival=0.005983040566669517, role=0.003406285248602807,
romance=0.005983040566669517, romeo=0.005983040566669517,
rope=0.005983040566669517, rose=0.004475735332066677, said=0.0,
screen=0.005983040566669517, screenwriter=0.005983040566669517,
she=0.002576755318066711, shows=0.005983040566669517,
side=0.01342720599620003, son=0.005983040566669517,
sondheim=0.005983040566669517, stage=0.011966081133339034,
star=0.002576755318066711, stephen=0.004475735332066677,
story=0.02393216226667807, street=0.004475735332066677,
streisand=0.011966081133339034, stripper=0.005983040566669517,
successfully=0.005983040566669517, such=0.0013259291596235405,
that=3.9167477939712394E-4, the=-0.005669006977208915, this=0.0,
three=0.003406285248602807, times=0.002576755318066711, to=-
7.086258721511144E-4, tony=0.003406285248602807,
training=0.005983040566669517, turned=0.005983040566669517,
turning=0.004475735332066677, version=0.005983040566669517,
war=0.005983040566669517, was=3.9167477939712394E-4,
way=0.003406285248602807, we=8.295299305360957E-4, were=0.002651858319247081,
west=0.010218855745808421, which=8.295299305360957E-4,
wholesale=0.005983040566669517, with=3.9167477939712394E-4,
won=0.005983040566669517, world=0.002576755318066711,
writer=0.005983040566669517, writing=0.005983040566669517,
wrote=0.005983040566669517, year=0.0018989800139999654,
york=0.008951470664133354, you=0.003406285248602807}
sort_words Func#####
sort_scores Func#####
make_query_vector Func#####

* current_word_setk : 1
[a]
The following synsets contain 'a' or a possible base form of that text:
[a, angstrom, angstrom unit, A, vitamin A, antiophthalmic factor, axerophthol, deoxyadenosine monophosphate, adenine, ampere, amp, letter a, type A, group A]

* current_word_setk : 1
[of]
[of]

* current_word_setk : 1
[for]
[for]

* current_word_setk : 1
[to]
[to]

* current_word_setk : 1
[has]
The following synsets contain 'has' or a possible base form of that text:
[has, hour angle, HA, have, have got, hold, feature, experience, receive, get, undergo, own, possess, let, consume, ingest, take in, take, throw, make, give, induce, stimulate, cause, accept, suffer, sustain, give birth, deliver, bear, birth]

* current_word_setk : 1
[at]
The following synsets contain 'at' or a possible base form of that text:
[at, astatine, At, atomic number 85]

* current_word_setk : 1
[by]
The following synsets contain 'by' or a possible base form of that text:
[by, past, aside, away]

* current_word_setk : 1
[from]
[from]

* current_word_setk : 1
[it]
The following synsets contain 'it' or a possible base form of that text:
[it, information technology, IT]

* current_word_setk : 1
[on]
The following synsets contain 'on' or a possible base form of that text:
[on, along]

* current_word_setk : 1
[said]
The following synsets contain 'said' or a possible base form of that text:
[said, state, say, tell, allege, aver, suppose, read, order, enjoin, pronounce, articulate, enounce, sound out, enunciate, aforesaid, aforementioned]

* current_word_setk : 1
[this]
[this]

* current_word_setk : 1
[that]
[that]

* current_word_setk : 1
[was]
The following synsets contain 'was' or a possible base form of that text:

[was, Washington, Evergreen State, WA, be, exist, equal, constitute, represent, make up, comprise, follow, embody, personify, live, cost]

* current_word_setk : 1

[with]

[with]

* current_word_setk : 1

[he]

The following synsets contain 'he' or a possible base form of that text:

[he, helium, He, atomic number 2]

* current_word_setk : 1

[most]

The following synsets contain 'most' or a possible base form of that text:

[most, to the highest degree, about, just about, almost, all but, nearly, near, nigh, virtually, well-nigh]

* current_word_setk : 1

[we]

[we]

* current_word_setk : 1

[which]

[which]

* current_word_setk : 1

[about]

The following synsets contain 'about' or a possible base form of that text:

[about, approximately, close to, just about, some, roughly, more or less, around, or so, almost, most, all but, nearly, near, nigh, virtually, well-nigh, astir]

* current_word_setk : 1

[can]

The following synsets contain 'can' or a possible base form of that text:

[can, tin, tin can, canful, can buoy, buttocks, nates, arse, butt, backside, bum, buns, fundament, hindquarters, hind end, keister, posterior, prat, rear, rear end, rump, stern, seat, tail, tail end, tooshie, tush, bottom, behind, derriere, fanny, ass, toilet, commode, crapper, pot, potty, stool, throne, lavatory, lav, john, privy, bathroom, put up, fire, give notice, dismiss, give the axe, send away, sack, force out, give the sack, terminate]

* current_word_setk : 1

[into]

[into]

* current_word_setk : 1

[other]

The following synsets contain 'other' or a possible base form of that text:

[other, early, former]

* current_word_setk : 1

[such]

The following synsets contain 'such' or a possible base form of that text:

[such, so much]

* current_word_setk : 1

[year]

The following synsets contain 'year' or a possible base form of that text:

[year, twelvemonth, yr, class]

* current_word_setk : 1

[new]

The following synsets contain 'new' or a possible base form of that text:

[new, recently, newly, freshly, fresh, unexampled, novel, raw, newfangled, young, New, Modern]

* current_word_setk : 1

[during]

[during]
 * current_word_setk : 1
 [earlier]
 The following synsets contain 'earlier' or a possible base form of that text:
 [earlier, early, before, sooner, in the first place, in the beginning, to begin with, originally, earliest, former, other]
 * current_word_setk : 1
 [later]
 The following synsets contain 'later' or a possible base form of that text:
 [later, late, subsequently, afterwards, afterward, after, later on, by and by, future, ulterior]
 * current_word_setk : 1
 [she]
 [she]
 * current_word_setk : 1
 [star]
 The following synsets contain 'star' or a possible base form of that text:
 [star, ace, adept, champion, sensation, maven, mavin, virtuoso, genius, hotshot, superstar, whiz, whizz, wizard, wiz, principal, lead, headliner, asterisk, star topology, leading, prima, starring, stellar]
 * current_word_setk : 1
 [times]
 The following synsets contain 'times' or a possible base form of that text:
 [times, multiplication, time, clip, clock time, fourth dimension, meter, metre, prison term, sentence, clock]
 * current_word_setk : 1
 [world]
 The following synsets contain 'world' or a possible base form of that text:
 [world, universe, existence, creation, cosmos, macrocosm, reality, domain, Earth, globe, worldly concern, earthly concern, earth, populace, public, human race, humanity, humankind, human beings, humans, mankind, man, global, planetary, worldwide, world-wide]
 * current_word_setk : 1
 [were]
 The following synsets contain 'were' or a possible base form of that text:
 [were, be, exist, equal, constitute, represent, make up, comprise, follow, embody, personify, live, cost]
 * current_word_setk : 1
 [as]
 The following synsets contain 'as' or a possible base form of that text:
 [as, arsenic, As, atomic number 33, American Samoa, Eastern Samoa, AS, angstrom, angstrom unit, A, vitamin A, antiophthalmic factor, axerophthol, deoxyadenosine monophosphate, adenine, ampere, amp, letter a, type A, group A, equally, every bit]
 * current_word_setk : 1
 [military]
 The following synsets contain 'military' or a possible base form of that text:
 [military, armed forces, armed services, military machine, war machine]
 * current_word_setk : 1
 [role]
 The following synsets contain 'role' or a possible base form of that text:
 [role, function, office, part, character, theatrical role, persona, purpose, use]
 * current_word_setk : 1
 [three]
 The following synsets contain 'three' or a possible base form of that text:

[three, 3, III, trio, threesome, tierce, leash, troika, triad, trine, trinity, ternary, ternion, triplet, tercet, terzetto, trey, deuce-ace, iii]

* current_word_setk : 1

[tony]

[tony]

* current_word_setk : 1

[way]

The following synsets contain 'way' or a possible base form of that text:

[way, manner, mode, style, fashion, means, agency, path, way of life, direction, room, elbow room, right smart]

* current_word_setk : 1

[you]

[you]

* current_word_setk : 1

[2008]

[2008]

* current_word_setk : 1

[aged]

The following synsets contain 'aged' or a possible base form of that text:

[aged, elderly, age, senesce, get on, mature, mature, older, senior, of age, ripened, cured]

* current_word_setk : 1

[agent]

The following synsets contain 'agent' or a possible base form of that text:

[agent, factor, broker, federal agent, agentive role]

* current_word_setk : 1

[award]

The following synsets contain 'award' or a possible base form of that text:

[award, awarding, accolade, honor, honour, laurels, prize, present, grant]

* current_word_setk : 1

[based]

The following synsets contain 'based' or a possible base form of that text:

[based, establish, base, ground, found, free-base, station, post, send, place, founded]

* current_word_setk : 1

[began]

The following synsets contain 'began' or a possible base form of that text:

[began, get down, begin, get, start out, start, set about, set out, commence, lead off]

* current_word_setk : 1

[former]

The following synsets contain 'former' or a possible base form of that text:

[former, erstwhile, old, onetime, quondam, sometime, late, previous, early, other]

* current_word_setk : 1

[get]

The following synsets contain 'get' or a possible base form of that text:

[get, acquire, become, go, let, have, receive, find, obtain, incur, arrive, come, bring, convey, fetch, experience, undergo, pay back, pay off, fix, make, induce, stimulate, cause, catch, capture, grow, develop, produce, contract, take, drive, aim, arrest, scam, buzz off, fuck off, bugger off, get under one's skin, draw, perplex, vex, stick, puzzle, mystify, baffle, beat, pose, bewilder, flummox, stupefy, nonplus, gravel, amaze, dumbfound, get down, begin, start out, start, set about, set out, commence, suffer, sustain, beget, engender, father, mother, sire, generate, bring forth]

* current_word_setk : 1

[hit]

The following synsets contain 'hit' or a possible base form of that text:
[hit, hitting, striking, smash, smasher, strike, bang, collision, impinge on, run into, collide with, reach, make, attain, arrive at, gain, shoot, pip, come to, stumble, score, tally, rack up, murder, slay, dispatch, bump off, off, polish off, remove]
* current_word_setk : 1
[i]

The following synsets contain 'i' or a possible base form of that text:
[i, iodine, iodine, I, atomic number 53, one, 1, ace, single, unity, letter i, ane]
* current_word_setk : 1
[introduced]

The following synsets contain 'introduced' or a possible base form of that text:
[introduced, introduce, present, acquaint, innovate, insert, enclose, inclose, stick in, put in, bring in, infix, enter, bring out, precede, preface, premise, inaugurate, usher in]
* current_word_setk : 1
[leading]

The following synsets contain 'leading' or a possible base form of that text:
[leading, lead, leadership, take, direct, conduct, guide, leave, result, head, run, go, pass, extend, top, contribute, conduce, precede, moderate, chair, prima, star, starring, stellar, preeminent, ahead, in the lead]
* current_word_setk : 1
[manhattan]

The following synsets contain 'manhattan' or a possible base form of that text:
[manhattan, Manhattan]
* current_word_setk : 1
[oscar-winning]
[oscar-winning]
* current_word_setk : 1
[point]

The following synsets contain 'point' or a possible base form of that text:
[point, degree, level, stage, detail, item, point in time, tip, peak, dot, spot, compass point, period, full stop, stop, full point, head, pointedness, gunpoint, power point, distributor point, breaker point, indicate, show, orient, charge, steer, maneuver, manoeuvre, manoeuvre, direct, guide, channelize, channelise, bespeak, betoken, signal, luff, target, aim, place, sharpen, taper, repoint]
* current_word_setk : 1
[radio]

The following synsets contain 'radio' or a possible base form of that text:
[radio, radiocommunication, wireless, radio receiver, receiving set, radio set, tuner]
* current_word_setk : 1
[recently]

The following synsets contain 'recently' or a possible base form of that text:
[recently, late, lately, of late, latterly, newly, freshly, fresh, new]
* current_word_setk : 1
[remains]

The following synsets contain 'remains' or a possible base form of that text:
[remains, cadaver, corpse, stiff, clay, stay, remain, rest, stay on, continue, persist]
* current_word_setk : 1
[rose]

The following synsets contain 'rose' or a possible base form of that text:
[rose, rosebush, blush wine, pink wine, rose wine, rosiness, rise, lift,
arise, move up, go up, come up, uprise, climb, get up, stand up, rear,
surface, rise up, heighten, originate, develop, spring up, grow, wax, mount,
turn out, jump, climb up, prove, rebel, ascend, resurrect, roseate,
rosaceous]
* current_word_setk : 1
[stephen]
The following synsets contain 'stephen' or a possible base form of that text:
[stephen, Stephen, Sir Leslie Stephen]
* current_word_setk : 1
[street]
The following synsets contain 'street' or a possible base form of that text:
[street]
* current_word_setk : 1
[turning]
The following synsets contain 'turning' or a possible base form of that text:
[turning, turn, become, change state, grow, move around, turn over, change by
reversal, reverse, plow, plough, release, twist, sprain, wrench, wrick, rick,
sour, ferment, work, flex, bend, deform, call on]
* current_word_setk : 1
[19-year-old]
[19-year-old]
* current_word_setk : 1
[1959]
[1959]
* current_word_setk : 1
[1962]
[1962]
* current_word_setk : 1
[1968]
[1968]
* current_word_setk : 1
[1977]
[1977]
* current_word_setk : 1
[1984]
[1984]
* current_word_setk : 1
[93]
The following synsets contain '93' or a possible base form of that text:
[93, ninety-three, xciii]
* current_word_setk : 1
[actress]
The following synsets contain 'actress' or a possible base form of that text:
[actress]
* current_word_setk : 1
[alfred]
The following synsets contain 'alfred' or a possible base form of that text:
[alfred, Alfred, AElfred, Alfred the Great]
* current_word_setk : 1
[another]
The following synsets contain 'another' or a possible base form of that text:
[another, some other]
* current_word_setk : 1
[attorney]

The following synsets contain 'attorney' or a possible base form of that text:

[attorney, lawyer]

* current_word_setk : 1

[author]

The following synsets contain 'author' or a possible base form of that text:

[author, writer, generator, source]

* current_word_setk : 1

[aux]

[aux]

* current_word_setk : 1

[baby!]

[baby!]

* current_word_setk : 1

[ballet]

The following synsets contain 'ballet' or a possible base form of that text:

[ballet, concert dance]

* current_word_setk : 1

[bernstein]

The following synsets contain 'bernstein' or a possible base form of that text:

[bernstein, Bernstein, Leonard Bernstein]

* current_word_setk : 1

[born]

The following synsets contain 'born' or a possible base form of that text:

[born, Born, Max Born, bear, give birth, deliver, birth, have, digest, endure, stick out, stomach, stand, tolerate, support, brook, abide, suffer, put up, turn out, take over, accept, assume, hold, carry, contain, yield, pay, wear, behave, acquit, deport, conduct, comport, have a bun in the oven, gestate, expect, natural, innate]

* current_word_setk : 1

[brooklyn]

The following synsets contain 'brooklyn' or a possible base form of that text:

[brooklyn, Brooklyn]

* current_word_setk : 1

[cage]

The following synsets contain 'cage' or a possible base form of that text:

[cage, coop, Cage, John Cage, John Milton Cage Jr., batting cage, cage in]

* current_word_setk : 1

[classic]

The following synsets contain 'classic' or a possible base form of that text:

[classic]

* current_word_setk : 1

[complications]

The following synsets contain 'complications' or a possible base form of that text:

[complications, complication, ramification, complicatedness, knottiness, tortuousness]

* current_word_setk : 1

[confirmed]

The following synsets contain 'confirmed' or a possible base form of that text:

[confirmed, confirm, corroborate, sustain, substantiate, support, affirm, reassert, habitual, inveterate]

* current_word_setk : 1

[dies]

The following synsets contain 'dies' or a possible base form of that text:
[dies, dice, die, dysprosium, Dy, atomic number 66, decease, perish, go, exit, pass away, expire, pass, kick the bucket, cash in one's chips, buy the farm, conk, give-up the ghost, drop dead, pop off, choke, croak, snuff it, fail, go bad, give way, give out, conk out, break, break down, die out, pall, become flat]

* current_word_setk : 1

[direct]

The following synsets contain 'direct' or a possible base form of that text:
[direct, target, aim, place, point, lead, take, conduct, guide, send, train, take aim, calculate, steer, maneuver, manoeuver, manoeuvre, head, channelize, channelise, address, mastermind, engineer, organize, organise, orchestrate, lineal, directly, straight, unmediated, verbatim]

* current_word_setk : 1

[directed]

The following synsets contain 'directed' or a possible base form of that text:

[directed, direct, target, aim, place, point, lead, take, conduct, guide, send, train, take aim, calculate, steer, maneuver, manoeuver, manoeuvre, head, channelize, channelise, address, mastermind, engineer, organize, organise, orchestrate]

* current_word_setk : 1

[directing]

The following synsets contain 'directing' or a possible base form of that text:

[directing, direct, target, aim, place, point, lead, take, conduct, guide, send, train, take aim, calculate, steer, maneuver, manoeuver, manoeuvre, head, channelize, channelise, address, mastermind, engineer, organize, organise, orchestrate, directional, directive, guiding]

* current_word_setk : 1

[featuring]

The following synsets contain 'featuring' or a possible base form of that text:

[featuring, have, feature, sport, boast]

* current_word_setk : 1

[films]

The following synsets contain 'films' or a possible base form of that text:
[films, movie, film, picture, moving picture, moving-picture show, motion picture, motion-picture show, picture show, pic, flick, cinema, celluloid, plastic film, photographic film, shoot, take]

* current_word_setk : 1

[folles]

[folles]

* current_word_setk : 1

[four]

The following synsets contain 'four' or a possible base form of that text:
[four, 4, IV, tetrad, quatern, quaternion, quaternary, quaternity, quartet, quadruplet, foursome, Little Joe, iv]

* current_word_setk : 1

[gangs]

The following synsets contain 'gangs' or a possible base form of that text:
[gangs, gang, pack, ring, mob, crowd, crew, bunch, work party, gang up]

* current_word_setk : 1

[hallelujah]

The following synsets contain 'hallelujah' or a possible base form of that text:

[hallelujah]

* current_word_setk : 1
[himself]
[himself]
* current_word_setk : 1
[hitchcock]
The following synsets contain 'hitchcock' or a possible base form of that text:
[hitchcock, Hitchcock, Alfred Hitchcock, Sir Alfred Hitchcock, Alfred Joseph Hitchcock]
* current_word_setk : 1
[home]
The following synsets contain 'home' or a possible base form of that text:
[home, place, dwelling, domicile, abode, habitation, dwelling house, nursing home, rest home, base, family, household, house, menage, home plate, home base, plate, interior, internal, national]
* current_word_setk : 1
[ii]
The following synsets contain 'ii' or a possible base form of that text:
[ii, two, 2, II, deuce]
* current_word_setk : 1
[juliet]
[juliet]
* current_word_setk : 1
[known]
The following synsets contain 'known' or a possible base form of that text:
[known, know, cognize, cognise, experience, live, acknowledge, recognize, recognise, roll in the hay, love, make out, make love, sleep with, get laid, have sex, do it, be intimate, have intercourse, have it away, have it off, screw, fuck, jazz, eff, hump, lie with, bed, have a go at it, bang, get it on, bonk]
* current_word_setk : 1
[la]
The following synsets contain 'la' or a possible base form of that text:
[la, lanthanum, La, atomic number 57, Louisiana, Pelican State, LA, lah]
* current_word_setk : 1
[latter]
The following synsets contain 'latter' or a possible base form of that text:
[latter]
* current_word_setk : 1
[lee]
The following synsets contain 'lee' or a possible base form of that text:
[lee, Lee, Spike Lee, Shelton Jackson Lee, Gypsy Rose Lee, Rose Louise Hovick, Bruce Lee, Lee Yuen Kam, Tsung Dao Lee, Richard Henry Lee, Henry Lee, Lighthorse Harry Lee, Robert E. Lee, Robert Edward Lee, lee side, leeward, downwind]
* current_word_setk : 1
[leonard]
The following synsets contain 'leonard' or a possible base form of that text:
[leonard, Leonard, Elmore Leonard, Elmore John Leonard, Dutch Leonard]
* current_word_setk : 1
[lupone]
[lupone]
* current_word_setk : 1
[lyrics]
The following synsets contain 'lyrics' or a possible base form of that text:
[lyrics, lyric, words, language, lyric poem]
* current_word_setk : 1

[memoirs]

The following synsets contain 'memoirs' or a possible base form of that text:

[memoirs, memoir]

* current_word_setk : 1

[movies]

The following synsets contain 'movies' or a possible base form of that text:

[movies, movie, film, picture, moving picture, moving-picture show, motion picture, motion-picture show, picture show, pic, flick]

* current_word_setk : 1

[music]

The following synsets contain 'music' or a possible base form of that text:

[music, euphony, medicine]

* current_word_setk : 1

[musicals]

The following synsets contain 'musicals' or a possible base form of that text:

[musicals, musical, musical comedy, musical theater]

* current_word_setk : 1

[patty]

The following synsets contain 'patty' or a possible base form of that text:

[patty, cake]

* current_word_setk : 1

[plans]

The following synsets contain 'plans' or a possible base form of that text:

[plans, plan, program, programme, design, architectural plan, be after, project, contrive]

* current_word_setk : 1

[playwright]

The following synsets contain 'playwright' or a possible base form of that text:

[playwright, dramatist]

* current_word_setk : 1

[pneumonia]

The following synsets contain 'pneumonia' or a possible base form of that text:

[pneumonia]

* current_word_setk : 1

[possibly]

The following synsets contain 'possibly' or a possible base form of that text:

[possibly, perchance, perhaps, maybe, mayhap, peradventure, potentially]

* current_word_setk : 1

[premiered]

The following synsets contain 'premiered' or a possible base form of that text:

[premiered, premier, premiere]

* current_word_setk : 1

[remembered]

The following synsets contain 'remembered' or a possible base form of that text:

[remembered, remember, retrieve, recall, call back, call up, recollect, think, think of, think back, commend, commemorate]

* current_word_setk : 1

[retold]

The following synsets contain 'retold' or a possible base form of that text:

[retold, recite, retell, fictionalize, fictionalise, repeat, reiterate, ingeminate, iterate, restate]

* current_word_setk : 1
[revivals]
The following synsets contain 'revivals' or a possible base form of that text:
[revivals, revival, resurgence, revitalization, revitalisation, revivification, revival meeting]
* current_word_setk : 1
[revived]
The following synsets contain 'revived' or a possible base form of that text:
[revived, resuscitate, revive, animate, recreate, reanimate, renovate, repair, quicken, vivify, revivify, resurrect, come to, reanimated]
* current_word_setk : 1
[rival]
The following synsets contain 'rival' or a possible base form of that text:
[rival, challenger, competitor, competition, contender, equal, touch, match]
* current_word_setk : 1
[romance]
The following synsets contain 'romance' or a possible base form of that text:
[romance, love affair, romanticism, Romance, Romance language, Latinian language, love story, woo, court, solicit, chat up, flirt, dally, butterfly, coquet, coquette, philander, mash, Latin]
* current_word_setk : 1
[romeo]
The following synsets contain 'romeo' or a possible base form of that text:
[romeo, Romeo]
* current_word_setk : 1
[rope]
The following synsets contain 'rope' or a possible base form of that text:
[rope, R-2, Mexican valium, rophy, roofy, roach, forget me drug, circle, lasso, leash]
* current_word_setk : 1
[screen]
The following synsets contain 'screen' or a possible base form of that text:
[screen, silver screen, projection screen, blind, CRT screen, cover, covert, concealment, sieve, screen door, test, screen out, sort, block out, riddle, shield]
* current_word_setk : 1
[screenwriter]
The following synsets contain 'screenwriter' or a possible base form of that text:
[screenwriter, film writer]
* current_word_setk : 1
[shows]
The following synsets contain 'shows' or a possible base form of that text:
[shows, show, display, appearance, demo, exhibit, present, demonstrate, prove, establish, shew, testify, bear witness, evidence, picture, depict, render, express, evince, indicate, point, reveal, show up, read, register, record, usher]
* current_word_setk : 1
[son]
The following synsets contain 'son' or a possible base form of that text:
[son, boy, Son, Word, Logos]
* current_word_setk : 1
[sondheim]
The following synsets contain 'sondheim' or a possible base form of that text:
[sondheim, Sondheim, Stephen Sondheim]

* current_word_setk : 1
[stripper]
The following synsets contain 'stripper' or a possible base form of that text:
[stripper, stemmer, sprigger, striptease artist, striptease, stripteaser, exotic dancer, ecdysiast, peeler, stripper well]

* current_word_setk : 1
[successfully]
The following synsets contain 'successfully' or a possible base form of that text:
[successfully, with success]

* current_word_setk : 1
[training]
The following synsets contain 'training' or a possible base form of that text:
[training, preparation, grooming, education, breeding, train, develop, prepare, educate, discipline, check, condition, groom, school, cultivate, civilize, civilise, aim, take, take aim, direct, coach, rail, trail]

* current_word_setk : 1
[turned]
The following synsets contain 'turned' or a possible base form of that text:
[turned, turn, become, change state, grow, move around, turn over, change by reversal, reverse, plow, plough, release, twist, sprain, wrench, wrick, rick, sour, ferment, work, flex, bend, deform, call on, off]

* current_word_setk : 1
[version]
The following synsets contain 'version' or a possible base form of that text:
[version, variant, variation, edition, adaptation, translation, interlingual rendition, rendering, interpretation, reading]

* current_word_setk : 1
[war]
The following synsets contain 'war' or a possible base form of that text:
[war, warfare, state of war]

* current_word_setk : 1
[wholesale]
The following synsets contain 'wholesale' or a possible base form of that text:
[wholesale, in large quantities, sweeping]

* current_word_setk : 1
[won]
The following synsets contain 'won' or a possible base form of that text:
[won, South Korean won, North Korean won, win, acquire, gain, advance, pull ahead, make headway, get ahead, gain ground, succeed, come through, bring home the bacon, deliver the goods]

* current_word_setk : 1
[writer]
The following synsets contain 'writer' or a possible base form of that text:
[writer, author]

* current_word_setk : 1
[writing]
The following synsets contain 'writing' or a possible base form of that text:
[writing, authorship, composition, penning, written material, piece of writing, committal to writing, write, compose, pen, indite, publish, drop a line, spell]

* current_word_setk : 1
[wrote]
The following synsets contain 'wrote' or a possible base form of that text:

[wrote, write, compose, pen, indite, publish, drop a line, spell]
 * current_word_setk : 1
 [director]
 The following synsets contain 'director' or a possible base form of that text:
 [director, manager, managing director, theater director, theatre director, film director, conductor, music director]
 * current_word_setk : 1
 [film]
 The following synsets contain 'film' or a possible base form of that text:
 [film, movie, picture, moving picture, moving-picture show, motion picture, motion-picture show, picture show, pic, flick, cinema, celluloid, plastic film, photographic film, shoot, take]
 * current_word_setk : 1
 [york]
 The following synsets contain 'york' or a possible base form of that text:
 [york, York, House of York]
 * current_word_setk : 1
 [west]
 The following synsets contain 'west' or a possible base form of that text:
 [west, West, Occident, due west, westward, W, western United States, Rebecca West, Dame Rebecca West, Cicily Isabel Fairfield, Mae West, Benjamin West]
 * current_word_setk : 1
 [his]
 The following synsets contain 'his' or a possible base form of that text:
 [his, hello, hullo, hi, howdy, how-do-you-do, Hawaii, Hawai'i, Aloha State, HI]
 * current_word_setk : 1
 [arthur]
 The following synsets contain 'arthur' or a possible base form of that text:
 [arthur, Arthur, Chester A. Arthur, Chester Alan Arthur, President Arthur, King Arthur]
 * current_word_setk : 1
 [barbra]
 [barbra]
 * current_word_setk : 1
 [best]
 The following synsets contain 'best' or a possible base form of that text:
 [best, topper, Best, C. H. Best, Charles Herbert Best, outdo, outflank, trump, scoop, good, better, well, easily, considerably, substantially, intimately, advantageously, comfortably, full, estimable, honorable, respectable, beneficial, just, upright, adept, expert, practiced, proficient, skillful, skilful, dear, near, dependable, safe, secure, right, ripe, effective, in effect, in force, serious, sound, salutary, unspoiled, unspoil]
 * current_word_setk : 1
 [book]
 The following synsets contain 'book' or a possible base form of that text:
 [book, volume, ledger, leger, account book, book of account, record, record book, script, playscript, rule book, Koran, Quran, al-Qur'an, Book, Bible, Christian Bible, Good Book, Holy Scripture, Holy Writ, Scripture, Word of God, Word, reserve, hold]
 * current_word_setk : 1
 [credits]
 The following synsets contain 'credits' or a possible base form of that text:
 [credits, recognition, credit, credit entry, deferred payment, course credit, citation, cite, acknowledgment, reference, mention, quotation, accredit]

```

* current_word_setk : 1
[died]
The following synsets contain 'died' or a possible base form of that text:
[died, die, decease, perish, go, exit, pass away, expire, pass, kick the
bucket, cash in one's chips, buy the farm, conk, give-up the ghost, drop
dead, pop off, choke, croak, snuff it, fail, go bad, give way, give out, conk
out, break, break down, die out, pall, become flat]
* current_word_setk : 1
[drama]
The following synsets contain 'drama' or a possible base form of that text:
[drama, play, dramatic play, dramatic event]
* current_word_setk : 1
[include]
The following synsets contain 'include' or a possible base form of that text:
[include, admit, let in]
* current_word_setk : 1
[musical]
The following synsets contain 'musical' or a possible base form of that text:
[musical, musical comedy, musical theater, melodious, melodic]
* current_word_setk : 1
[stage]
The following synsets contain 'stage' or a possible base form of that text:
[stage, phase, degree, level, point, stagecoach, leg, microscope stage,
present, represent, arrange]
* current_word_setk : 1
[streisand]
The following synsets contain 'streisand' or a possible base form of that
text:
[streisand, Streisand, Barbra Streisand, Barbra Joan Streisand]
* current_word_setk : 1
[broadway]
The following synsets contain 'broadway' or a possible base form of that
text:
[broadway, Broadway, Great White Way]
* current_word_setk : 1
[side]
The following synsets contain 'side' or a possible base form of that text:
[side, face, side of meat, position, slope, incline, English, pull, root, go
with]
* current_word_setk : 1
[gypsy]
The following synsets contain 'gypsy' or a possible base form of that text:
[gypsy, itinerant, gipsy, Gypsy, Gipsy, Romany, Rommany, Romani, Roma,
Bohemian]
* current_word_setk : 1
[laurents]
[laurents]
* current_word_setk : 1
[story]
The following synsets contain 'story' or a possible base form of that text:
[story, narrative, narration, tale, floor, level, storey, history, account,
chronicle, report, news report, write up, fib, tarradiddle, taradiddle]
* * * Just to check : 165

former      in both 5    7
other in both 5    10
former      in both 7    10

```

class	in both	1	7
class	in both	1	10
class	in both	1	11
yr	in both	7	10
yr	in both	7	11
year	in both	10	11
new	in both	1	7
recently	in both	1	9
new	in both	1	11
recently	in both	7	9
new	in both	7	11
recently	in both	9	11
other	in both	5	7
other	in both	5	10
other	in both	5	11
other	in both	7	10
former	in both	7	11
other	in both	10	11
lead	in both	9	11
man	in both	2	3
man	in both	2	4
man	in both	3	4
were	in both	4	6
3	in both	1	9
way	in both	4	6
way	in both	4	8
way	in both	4	10
way	in both	6	8
way	in both	6	10
way	in both	8	10
age	in both	1	2
age	in both	1	5
age	in both	1	10
age	in both	2	5
age	in both	2	10
age	in both	5	10
get	in both	3	10
other	in both	5	6
other	in both	5	7
other	in both	5	10
late	in both	6	7
late	in both	6	10
old	in both	7	10
let	in both	4	10
hit	in both	4	6
direct	in both	2	5
direct	in both	2	9
direct	in both	2	10
direct	in both	2	11
direct	in both	5	9
direct	in both	5	10
direct	in both	5	11
direct	in both	9	10
direct	in both	9	11
direct	in both	10	11
direct	in both	1	2
stage	in both	1	4

direct	in both 1	5
show	in both 1	6
direct	in both 1	9
stage	in both 1	10
direct	in both 1	11
direct	in both 2	4
direct	in both 2	5
direct	in both 2	6
direct	in both 2	9
direct	in both 2	10
direct	in both 2	11
direct	in both 4	5
show	in both 4	6
direct	in both 4	9
stage	in both 4	10
direct	in both 4	11
direct	in both 5	6
direct	in both 5	9
direct	in both 5	10
direct	in both 5	11
direct	in both 6	9
show	in both 6	10
direct	in both 6	11
direct	in both 9	10
direct	in both 9	11
direct	in both 10	11
new	in both 1	6
new	in both 1	7
new	in both 1	9
new	in both 1	11
new	in both 6	7
late	in both 6	9
new	in both 6	11
new	in both 7	9
new	in both 7	11
new	in both 9	11
turn	in both 4	6
writer	in both 1	2
writer	in both 1	5
writer	in both 2	5
die	in both 0	1
die	in both 0	2
die	in both 1	2
train	in both 2	3
point	in both 2	4
direct	in both 2	5
direct	in both 2	9
direct	in both 2	10
direct	in both 2	11
train	in both 3	4
train	in both 3	5
train	in both 3	9
train	in both 3	10
train	in both 3	11
point	in both 4	5
point	in both 4	9
point	in both 4	10

```

point in both 4    11
direct      in both 5    9
direct      in both 5   10
direct      in both 5   11
direct      in both 9   10
direct      in both 9   11
direct      in both 10  11
train in both 2    3
point in both 2    4
direct      in both 2    5
direct      in both 2    9
direct      in both 2   10
direct      in both 2   11
train in both 3    4
train in both 3    5
train in both 3    9
train in both 3   10
train in both 3   11
point in both 4    5
point in both 4    9
point in both 4   10
point in both 4   11
direct      in both 5    9
direct      in both 5   10
direct      in both 5   11
direct      in both 9   10
direct      in both 9   11
direct      in both 10  11
train in both 2    3
point in both 2    4
direct      in both 2    5
direct      in both 2    9
direct      in both 2   10
direct      in both 2   11
train in both 3    4
train in both 3    5
train in both 3    9
train in both 3   10
train in both 3   11
point in both 4    5
point in both 4    9
point in both 4   10
point in both 4   11
direct      in both 5    9
direct      in both 5   10
direct      in both 5   11
direct      in both 9   10
direct      in both 9   11
direct      in both 10  11
film  in both 3    4
movie in both 3    6
film  in both 3   11
film  in both 4    6
film  in both 4   11
film  in both 6   11
iv    in both 5    7
4     in both 5    8

```


iv	in both	5	9
iv	in both	7	8
iv	in both	7	9
iv	in both	8	9
ring	in both	3	7
base	in both	2	8
2	in both	3	9
2	in both	3	10
2	in both	9	10
la	in both	0	1
la	in both	0	5
la	in both	0	6
la	in both	0	8
la	in both	0	9
la	in both	0	11
la	in both	1	5
la	in both	1	6
la	in both	1	8
la	in both	1	9
la	in both	1	11
la	in both	5	6
la	in both	5	8
la	in both	5	9
la	in both	5	11
la	in both	6	8
la	in both	6	9
la	in both	6	11
la	in both	8	9
la	in both	8	11
la	in both	9	11
film	in both	3	4
film	in both	3	6
film	in both	3	11
film	in both	4	6
film	in both	4	11
film	in both	6	11
music	in both	1	5
music	in both	1	7
music	in both	1	10
music	in both	5	7
music	in both	5	10
music	in both	7	10
musical	in both	1	5
musical	in both	1	10
musical	in both	5	10
screen	in both	2	4
point	in both	4	6
son	in both	3	7
direct	in both	2	3
direct	in both	2	5
direct	in both	2	9
direct	in both	2	10
direct	in both	2	11
direct	in both	3	5
direct	in both	3	9
direct	in both	3	10
direct	in both	3	11

direct	in both 5	9
direct	in both 5	10
direct	in both 5	11
direct	in both 9	10
direct	in both 9	11
direct	in both 10	11
turn	in both 4	6
war	in both 3	5
win	in both 5	11
writer	in both 1	2
author	in both 1	5
author	in both 2	5
write	in both 1	2
write	in both 1	6
write	in both 2	6
write	in both 1	2
write	in both 1	3
write	in both 2	3
director	in both 2	10
film	in both 3	4
movie	in both 3	6
film	in both 3	11
movie	in both 4	6
film	in both 4	11
movie	in both 6	11
york	in both 1	7
west	in both 0	1
west	in both 0	6
west	in both 1	6
his	in both 2	4
hi	in both 2	6
hi	in both 2	9
his	in both 2	10
his	in both 2	11
hi	in both 4	6
hi	in both 4	9
his	in both 4	10
his	in both 4	11
hi	in both 6	9
hi	in both 6	10
hi	in both 6	11
hi	in both 9	10
hi	in both 9	11
his	in both 10	11
arthur	in both 0	1
barbra	in both 4	10
right	in both 0	6
right	in both 0	8
right	in both 0	10
full	in both 6	8
best	in both 6	10
full	in both 8	10
book	in both 5	6
credits	in both 4	10
die	in both 0	1
die	in both 0	2
died	in both 1	2

play in both 0 4
 play in both 0 7
 drama in both 4 7
 include in both 4 10
 musical in both 1 5
 musical in both 1 10
 musical in both 5 10
 point in both 1 4
 stage in both 1 10
 point in both 4 10
 streisand in both 4 10
 broadway in both 6 8
 broadway in both 6 10
 broadway in both 8 10
 side in both 0 1
 side in both 0 6
 side in both 1 6
 gypsy in both 1 6
 gypsy in both 1 8
 gypsy in both 1 11
 gypsy in both 6 8
 gypsy in both 6 11
 gypsy in both 8 11
 laurents in both 0 1
 laurents in both 0 5
 laurents in both 0 9
 laurents in both 1 5
 laurents in both 1 9
 laurents in both 5 9
 story in both 0 1
 story in both 0 6
 story in both 0 7
 story in both 1 6
 story in both 1 7
 story in both 6 7

0	12	5	3	4	5	9	5	4	4	4	4
=59.0 8.428571428571429											
12	0	14	7	9	17	17	12	8	11	16	12
=135.0 5.869565217391305											
5	14	0	15	13	17	11	8	9	15	16	15
=138.0 8.117647058823529											
3	7	15	0	15	13	14	13	9	13	16	15
=133.0 6.65											
4	9	13	15	0	10	20	8	7	11	21	14
=132.0 5.5											
5	17	17	13	10	0	13	16	11	16	23	18
=159.0 5.129032258064516											
9	17	11	14	20	13	0	14	13	11	19	15
=156.0 6.5											

5	12	8	13	8	16	14	0	11	12	16	12
=127.0		4.535714285714286									
4	8	9	9	7	11	13	11	0	10	11	10
=103.0		4.291666666666667									
4	11	15	13	11	16	11	12	10	0	14	18
=135.0		7.105263157894737									
4	16	16	16	21	23	19	16	11	14	0	16
=172.0		5.548387096774194									
4	12	15	15	14	18	15	12	10	18	16	0
=149.0		7.095238095238095									

```

sentence_scores_tf_idf_normal scores:          [-0.37842217395538147,
1.9093784486230523, -1.1689800800584613, -1.2504521044480517,
0.3376416947862209, 0.4491992846144839, 0.8818115473910257,
0.17103951942010928, -0.3700769624686484, -1.016417479296216,
1.1906506170393563, -0.7553723116474882]
sentence_scores_tf_idf_normal scores:          [-0.37842217395538147,
1.9093784486230523, -1.1689800800584613, -1.2504521044480517,
0.3376416947862209, 0.4491992846144839, 0.8818115473910257,
0.17103951942010928, -0.3700769624686484, -1.016417479296216,
1.1906506170393563, -0.7553723116474882]
sentence_vectors_similarity_query_scores_tf_idf_normal scores:          [-
0.28867513459481287, -0.28867513459481287, -0.28867513459481287, -
0.28867513459481287, -0.28867513459481287, -0.28867513459481287, -
0.28867513459481287, -0.28867513459481287, -0.28867513459481287, -
0.28867513459481287, -0.28867513459481287, 3.1754264805429417]
sentence_vectors_similarity_average_scores_tf_idf_normal scores:
[1.6283527308210426, 1.9795330331897134, -0.7646950666454064, -
1.2118531806134591, -0.9298839702601773, 0.13238166230723383,
0.7860295949050827, -0.04199955316737382, -0.547469273120486,
0.011552226391152028, -0.6476034831197174, -0.39434472068760507]
lexical_chain_normal scores:          [-2.533395494082398, 0.06262325940428433,
0.1650976838840218, -0.005693023582207314, -0.03985116507545314,
0.8824186552421841, 0.7799442307624466, -0.21064187254168226, -
1.030437268379582, 0.06262325940428433, 1.3264744946543798,
0.5408372403097259]
headerlines_1_normal scores:          [2.4085462193013756, 1.8418294618186988,
-0.42503756811200744, -0.42503756811200744, -0.42503756811200744, -
0.42503756811200744, -0.42503756811200744, -0.42503756811200744, -
0.42503756811200744, -0.42503756811200744, -0.42503756811200744, -
0.42503756811200744]
stigmas_1_normal scores:          [0.28867513459481314,
0.28867513459481314, -3.1754264805429413, 0.28867513459481314,
0.28867513459481314, 0.28867513459481314, 0.28867513459481314,
0.28867513459481314, 0.28867513459481314, 0.28867513459481314,
0.28867513459481314, 0.28867513459481314]
cue_1_normal scores:          [-0.2886751345948129, -0.2886751345948129,
3.1754264805429417, -0.2886751345948129, -0.2886751345948129, -
0.2886751345948129, -0.2886751345948129, -0.2886751345948129, -
0.2886751345948129, -0.2886751345948129, -0.2886751345948129, -
0.2886751345948129]

```

```

position_1_normal scores:      [1.9002239072636002, 1.7036490203052967,
0.5241996985554758, 0.3276248115971723, 0.13104992463886886, -
0.06552496231943461, -0.2620998492777381, -0.45867473623604155, -
0.655249623194345, -0.8518245101526485, -1.048399397110952, -
1.2449742840692555]
#####Scores Integration Module Func#####
[2.0, 3.0, 2.0, 2.0, 2.0, 2.0, 2.0, 3.0, 3.0, 2.0, 4.0]
*&*&*****
    final sentence scores based on scores : [15.831650063202948,
7.08480888217564, -3.680656369981719, -5.669222722504371, -
2.5871376471906653, 1.015323777361047, 2.3931706582754426, -
3.255328561294471, -7.577816417528922, -6.156058057467957, -
1.1222554741932718, 3.7235218691463086]
    final rank of sentences:      [0, 1, 11, 6, 5, 10, 4, 7, 2, 3, 9, 8]
[west side story playwright arthur laurents dies , arthur laurents, writer of
such classic stage musicals as west side story and gypsy, has died in new
york aged 93.,earlier this year the oscar-winning actress confirmed she plans
to star in and possibly direct a new film version of gypsy.]

```

4) Score Integration Module:

I have implemented two different ways for integrating the score of the sentence score modules:

i. Score Based

In this method, the scores of all the sentence scoring modules are normalized so that they are brought into the same scale, then they are combined together according to the weight assigned to each scoring module. The weights are defined based on the result of running the algorithm on the development data and trial and error in order to improve the summarization results.

For example, after the first set of experiment on the data I realized that in 2 of the documents, the order of the sentences is not kept and that can affect the coherency. For example in the second document above, sentence number 0 would be printed out after the sentence number 11. Therefore I increased the weight on the positional ordering of the sentences.

For example, in order to be able to get different results according to different query words, I needed to increase the weight of the similarity to the query.

For example , in order to be able to pick up different sentences from documents that have multiple headers and topics, I had to increase the weight of the position score module, so that the first lines of the different paragraphs would be selected.

The current weight vector is: [2.0, 3.0, 2.0, 2.0, 3.0, 3.0, 2.0, 4.0]

For the following score modules:

```

entence_scores_tf_idf_normal scores
scoressentence_vectors_similarity_query_scores_tf_idf_normal scores

```

```

scoresentence_vectors_similarity_average_scores_tf_idf_normal scores
scoreslexical_chain_normal scores
headerlines_1_normal scores
stigmas_1_normal scores
cue_1_normal scores
position_1_normal scores

```

```

private static ArrayList<String> score_module_2(ArrayList<String> inputlines)
{
    System.out.println("#####"+"Rank Based Integration
Func"+"#####");

    TreeMap<String, Integer>
word_count_1=term_freq_count(inputlines_1);
    TreeMap<String, Integer>
word_count_2=term_freq_count(inputlines_2);
    TreeMap<String, Integer>
word_count_3=term_freq_count(inputlines_3);
    TreeMap<String, Integer>
word_count_4=term_freq_count(inputlines_4);
    TreeMap<String, Integer>
word_count_5=term_freq_count(inputlines_5);
    TreeMap<String, Integer>
word_count_6=term_freq_count(inputlines_6);
    TreeMap<String, Integer>
word_count_7=term_freq_count(inputlines_7);
    TreeMap<String, Integer>
word_count_8=term_freq_count(inputlines_8);
    TreeMap<String, Integer>
word_count_9=term_freq_count(inputlines_9);
    TreeMap<String, Integer>
word_count_10=term_freq_count(inputlines_10);

    ArrayList<TreeMap<String, Integer>> all=new
ArrayList<TreeMap<String, Integer>>();
    all.add(word_count_1);
    all.add(word_count_2);
    all.add(word_count_3);
    all.add(word_count_4);
    all.add(word_count_5);
    all.add(word_count_6);
    all.add(word_count_7);
    all.add(word_count_8);
    all.add(word_count_9);
    all.add(word_count_10);

    int doc_size_1=doc_size(inputlines_1);
    length_1=sentence_length(inputlines_1);

    TreeMap<String, Double> word_tdf_1=
term_tdf(word_count_1,doc_size_1);
    TreeMap<String, Double> word_idf_1= term_idf(word_count_1, all);

```

```

        TreeMap<String, Double> word_tf_idf_1= term_tdf_idf(word_tdf_1,
word_idf_1);

        System.out.println("tf: "+word_tdf_1.toString());
        System.out.println("idf: "+word_idf_1.toString());
        System.out.println("tf-idf: "+word_tf_idf_1.toString());

        /////////// tf-idf
        ArrayList<String> sorted_words_tf_idf=sort_words(word_tf_idf_1);
        ArrayList<Double>
sorted_scores_tf_idf=sort_scores(word_tf_idf_1);

        ArrayList<Double> query_vector_tf_idf= make_query_vector(
sorted_words_tf_idf, sorted_scores_tf_idf, inputlines);
        ArrayList<Double> average_vector_tf_idf= make_average_vector(
sorted_words_tf_idf, sorted_scores_tf_idf, inputlines);
        ArrayList<ArrayList<Double>> sentence_vectors_tf_idf=
make_sentence_vectors( sorted_words_tf_idf, sorted_scores_tf_idf,
inputlines);

        ArrayList<Double>
sentence_scores_tf_idf=calculate_sentence_score(sentence_vectors_tf_idf);
        ArrayList<Double>
sentence_scores_tf_idf_normal=MathFuncs.normalizedArray(sentence_scores_tf_id
f);

        ArrayList<Double>
sentence_vectors_similarity_query_scores_tf_idf=calculate_similarity_score(se
ntence_vectors_tf_idf, query_vector_tf_idf);
        ArrayList<Double>
sentence_vectors_similarity_query_scores_tf_idf_normal=MathFuncs.normalizedAr
ray(sentence_vectors_similarity_query_scores_tf_idf);

        ArrayList<Double>
sentence_vectors_similarity_average_scores_tf_idf=calculate_similarity_score(
sentence_vectors_tf_idf, average_vector_tf_idf);
        ArrayList<Double>
sentence_vectors_similarity_average_scores_tf_idf_normal=MathFuncs.normalized
Array(sentence_vectors_similarity_average_scores_tf_idf);

        System.out.println("\n1: sentence tf idf
score:"+sentence_scores_tf_idf.toString());
        System.out.println("2: query
sim:"+sentence_vectors_similarity_query_scores_tf_idf.toString());
        System.out.println("3: average
sim:"+sentence_vectors_similarity_average_scores_tf_idf.toString());

        /////////// Lexicon Chain

        ArrayList<Double> lexical_chain=
score_module_Lexical_Chain(inputlines, sorted_words_tf_idf, length_1 );
        ArrayList<Double>
lexical_chain_normal=MathFuncs.normalizedArray(lexical_chain);

        ArrayList<Double>
headerlines_1_normal=MathFuncs.normalizedArray(headerlines_1);

```

```

        ArrayList<Double>
stigmas_1_normal=MathFuncs.normalizedArray(stigmas_1);
        ArrayList<Double> cue_1_normal=MathFuncs.normalizedArray(cue_1);
        ArrayList<Double>
position_1_normal=MathFuncs.normalizedArray(position_1);

        //////////
        ArrayList<ArrayList<Double>> all_scores= new
ArrayList<ArrayList<Double>>();

        all_scores.add(sentence_scores_tf_idf_normal);
        System.out.println("sentence_scores_tf_idf_normal
scores:\t\t"+sentence_scores_tf_idf_normal.toString());
        System.out.println("sentence_scores_tf_idf_normal
scores:\t\t"+sentence_scores_tf_idf_normal.toString());

        all_scores.add(sentence_vectors_similarity_query_scores_tf_idf_normal);

        System.out.println("sentence_vectors_similarity_query_scores_tf_idf_nor
mal
scores:\t\t"+sentence_vectors_similarity_query_scores_tf_idf_normal.toString(
));

        all_scores.add(sentence_vectors_similarity_average_scores_tf_idf_normal
);

        System.out.println("sentence_vectors_similarity_average_scores_tf_idf_n
ormal
scores:\t\t"+sentence_vectors_similarity_average_scores_tf_idf_normal.toStrin
g());

        all_scores.add(lexical_chain_normal);
        System.out.println("lexical_chain_normal
scores:\t\t"+lexical_chain_normal.toString());

        all_scores.add(headerlines_1_normal);
        System.out.println("headerlines_1_normal
scores:\t\t"+headerlines_1_normal.toString());

        all_scores.add(stigmas_1_normal);
        System.out.println("stigmas_1_normal
scores:\t\t"+stigmas_1_normal.toString());

        all_scores.add(cue_1_normal);
        System.out.println("cue_1_normal
scores:\t\t"+cue_1_normal.toString());

        all_scores.add(position_1_normal);
        System.out.println("position_1_normal
scores:\t\t"+position_1_normal.toString());

        ArrayList<Double> all_score_weights= new ArrayList<Double>();
        all_score_weights.add(2.0);
        all_score_weights.add(3.0);
        all_score_weights.add(2.0);

```



```

        all_score_weights.add(2.0);
        all_score_weights.add(2.0);
        all_score_weights.add(2.0);
        all_score_weights.add(2.0);
        all_score_weights.add(3.0);
        all_score_weights.add(3.0);
        all_score_weights.add(2.0);
        all_score_weights.add(4.0);

        ArrayList<Integer> final_rank=
score_vote(all_scores,all_score_weights);

        int number_of_lines_in_summary=
(int)((double)inputlines.size()*0.3);
        ArrayList<String> summary= convert_rank_string(inputlines,
number_of_lines_in_summary, final_rank);

        return summary;
    }

```

ii. Rank Based

In this method, the sentences are ranked according to the measure score that is the result of running each score module on the set of sentences. After ranking the sentences based on each score module, I use a voting schema to come up with the final ranking of the sentences. The voting takes into account some weight for each scoring module. For example sentences one can be ranked 1st by module1, be ranked 3rd by module2 and be ranked 4th by module3, the rank numbers are weighted based on function and a new rank number is derived. After deriving the new rank number for all sentences, they are sorted again based on the new ranking and the top 0.3 sentences are printed out as result.

The weights were also selected based on the results of running the algorithm on development text files.

The current weight numbers are:

[2, 2, 2, 2, 6, 2, 2, 2, 3, 2, 6]

That correspond to the following score modules:

```

sentence tf:
query sim tf:
average sim tf:
sentence tf idf:
query sim tf idf:
average sim tf idf:
lexical chain:
headline:
stigma:
cue:
position:

```

```

private static ArrayList<String> score_module_1(ArrayList<String> inputlines)
{

    System.out.println("#####"+ " Score Integration Based
On Scores Func"+"#####");
}

```

```

        TreeMap<String, Integer>
word_count_1=term_freq_count(inputlines_1);
        TreeMap<String, Integer>
word_count_2=term_freq_count(inputlines_2);
        TreeMap<String, Integer>
word_count_3=term_freq_count(inputlines_3);
        TreeMap<String, Integer>
word_count_4=term_freq_count(inputlines_4);
        TreeMap<String, Integer>
word_count_5=term_freq_count(inputlines_5);
        TreeMap<String, Integer>
word_count_6=term_freq_count(inputlines_6);
        TreeMap<String, Integer>
word_count_7=term_freq_count(inputlines_7);
        TreeMap<String, Integer>
word_count_8=term_freq_count(inputlines_8);
        TreeMap<String, Integer>
word_count_9=term_freq_count(inputlines_9);
        TreeMap<String, Integer>
word_count_10=term_freq_count(inputlines_10);

        ArrayList<TreeMap<String, Integer>> all=new
ArrayList<TreeMap<String, Integer>>();
        all.add(word_count_1);
        all.add(word_count_2);
        all.add(word_count_3);
        all.add(word_count_4);
        all.add(word_count_5);
        all.add(word_count_6);
        all.add(word_count_7);
        all.add(word_count_8);
        all.add(word_count_9);
        all.add(word_count_10);

        int doc_size_1=doc_size(inputlines_1);
        length_1=sentence_length(inputlines_1);
        /*int doc_size_2=doc_size(inputlines_2);
        int doc_size_3=doc_size(inputlines_3);
        int doc_size_4=doc_size(inputlines_4);
        int doc_size_5=doc_size(inputlines_5);
        int doc_size_6=doc_size(inputlines_6);
        int doc_size_7=doc_size(inputlines_7);
        int doc_size_8=doc_size(inputlines_8);
        int doc_size_9=doc_size(inputlines_9);
        int doc_size_10=doc_size(inputlines_10);
*/
        TreeMap<String, Double> word_tdf_1=
term_tdf(word_count_1,doc_size_1);
        /*TreeMap<String, Double> word_tdf_2=
term_tdf(word_count_2,doc_size_2);
        TreeMap<String, Double> word_tdf_3=
term_tdf(word_count_3,doc_size_3);
        TreeMap<String, Double> word_tdf_4=
term_tdf(word_count_4,doc_size_4);
        TreeMap<String, Double> word_tdf_5=
term_tdf(word_count_5,doc_size_5);

```

```

        TreeMap<String, Double> word_tdf_6=
term_tdf(word_count_6,doc_size_6);
        TreeMap<String, Double> word_tdf_7=
term_tdf(word_count_7,doc_size_7);
        TreeMap<String, Double> word_tdf_8=
term_tdf(word_count_8,doc_size_8);
        TreeMap<String, Double> word_tdf_9=
term_tdf(word_count_9,doc_size_9);
        TreeMap<String, Double> word_tdf_10=
term_tdf(word_count_10,doc_size_10);
*/

        TreeMap<String, Double> word_idf_1= term_idf(word_count_1, all);
/*TreeMap<String, Double> word_idf_2= term_idf(word_count_2,
all);

        TreeMap<String, Double> word_idf_3= term_idf(word_count_3, all);
TreeMap<String, Double> word_idf_4= term_idf(word_count_4, all);
TreeMap<String, Double> word_idf_5= term_idf(word_count_5, all);
TreeMap<String, Double> word_idf_6= term_idf(word_count_6, all);
TreeMap<String, Double> word_idf_7= term_idf(word_count_7, all);
TreeMap<String, Double> word_idf_8= term_idf(word_count_8, all);
TreeMap<String, Double> word_idf_9= term_idf(word_count_9, all);
TreeMap<String, Double> word_idf_10= term_idf(word_count_10,
all);
*/

        TreeMap<String, Double> word_tf_idf_1= term_tdf_idf(word_tdf_1,
word_idf_1);
/*TreeMap<String, Double> word_tf_idf_2= term_tdf_idf(word_tdf_2,
word_idf_2);
        TreeMap<String, Double> word_tf_idf_3= term_tdf_idf(word_tdf_3,
word_idf_3);
        TreeMap<String, Double> word_tf_idf_4= term_tdf_idf(word_tdf_4,
word_idf_4);
        TreeMap<String, Double> word_tf_idf_5= term_tdf_idf(word_tdf_5,
word_idf_5);
        TreeMap<String, Double> word_tf_idf_6= term_tdf_idf(word_tdf_6,
word_idf_6);
        TreeMap<String, Double> word_tf_idf_7= term_tdf_idf(word_tdf_7,
word_idf_7);
        TreeMap<String, Double> word_tf_idf_8= term_tdf_idf(word_tdf_8,
word_idf_8);
        TreeMap<String, Double> word_tf_idf_9= term_tdf_idf(word_tdf_9,
word_idf_9);
        TreeMap<String, Double> word_tf_idf_10= term_tdf_idf(word_tdf_10,
word_idf_10);
*/

        System.out.println("tf: "+word_tdf_1.toString());
        System.out.println("idf: "+word_idf_1.toString());
        System.out.println("tf-idf: "+word_tf_idf_1.toString());

        /////////// tf
        ArrayList<String> sorted_words=sort_words(word_tdf_1);
        ArrayList<Double> sorted_scores=sort_scores(word_tdf_1);

        ArrayList<Double> query_vector= make_query_vector( sorted_words,
sorted_scores, inputlines);
        ArrayList<Double> average_vector= make_average_vector(
sorted_words, sorted_scores, inputlines);

```

```

        ArrayList<ArrayList<Double>> sentence_vectors=
make_sentence_vectors( sorted_words, sorted_scores, inputlines);

        ArrayList<Double>
sentence_scores=calculate_sentence_score(sentence_vectors);
        ArrayList<Integer> rank_1= rank_by_score(sentence_scores);
        ArrayList<Double>
sentence_vectors_similarity_query_scores=calculate_similarity_score(sentence_
vectors, query_vector);
        ArrayList<Integer> rank_2=
rank_by_score(sentence_vectors_similarity_query_scores);
        ArrayList<Double>
sentence_vectors_similarity_average_scores=calculate_similarity_score(sentenc
e_vectors, average_vector);
        ArrayList<Integer> rank_3=
rank_by_score(sentence_vectors_similarity_average_scores);

        System.out.println("\n1: sentence tf
score:"+sentence_scores.toString());
        System.out.println("2: query
sim:"+sentence_vectors_similarity_query_scores.toString());
        System.out.println("3: average
sim:"+sentence_vectors_similarity_average_scores.toString());

        /////////// tf-idf
        ArrayList<String> sorted_words_tf_idf=sort_words(word_tf_idf_1);
        ArrayList<Double>
sorted_scores_tf_idf=sort_scores(word_tf_idf_1);

        ArrayList<Double> query_vector_tf_idf= make_query_vector(
sorted_words_tf_idf, sorted_scores_tf_idf, inputlines);
        ArrayList<Double> average_vector_tf_idf= make_average_vector(
sorted_words_tf_idf, sorted_scores_tf_idf, inputlines);
        ArrayList<ArrayList<Double>> sentence_vectors_tf_idf=
make_sentence_vectors( sorted_words_tf_idf, sorted_scores_tf_idf,
inputlines);

        ArrayList<Double>
sentence_scores_tf_idf=calculate_sentence_score(sentence_vectors_tf_idf);
        ArrayList<Integer> rank_4= rank_by_score(sentence_scores_tf_idf);
        ArrayList<Double>
sentence_vectors_similarity_query_scores_tf_idf=calculate_similarity_score(se
ntence_vectors_tf_idf, query_vector_tf_idf);
        ArrayList<Integer> rank_5=
rank_by_score(sentence_vectors_similarity_query_scores_tf_idf);
        ArrayList<Double>
sentence_vectors_similarity_average_scores_tf_idf=calculate_similarity_score(
sentence_vectors_tf_idf, average_vector_tf_idf);
        ArrayList<Integer> rank_6=
rank_by_score(sentence_vectors_similarity_average_scores_tf_idf);

        System.out.println("\n1: sentence tf idf
score:"+sentence_scores_tf_idf.toString());
        System.out.println("2: query
sim:"+sentence_vectors_similarity_query_scores_tf_idf.toString());
        System.out.println("3: average
sim:"+sentence_vectors_similarity_average_scores_tf_idf.toString());

```

```

////////// Lexicon Chain

ArrayList<Double> lexical_chain=
score_module_Lexical_Chain(inputlines, sorted words tf idf,length 1 );
ArrayList<Integer> rank_7= rank_by_score(lexical_chain);

ArrayList<Integer> rank_8= rank_by_score(headerlines 1);
ArrayList<Integer> rank_9= rank_by_score(stigmas 1);
ArrayList<Integer> rank_10= rank_by_score(cue 1);
ArrayList<Integer> rank_11= rank_by_score(position 1);

//////////
System.out.println("sentence tf:\t\t"+rank_1.toString());
System.out.println("query sim tf:\t\t"+rank_2.toString());
System.out.println("average sim tf:\t\t"+rank_3.toString());
System.out.println("sentence tf idf:\t\t"+rank_4.toString());
System.out.println("query sim tf idf:\t\t"+rank_5.toString());
System.out.println("average sim tf idf:\t\t"+rank_6.toString());
System.out.println("lexical chain:\t\t"+rank_7.toString());
System.out.println("headline:\t\t"+rank_8.toString());
System.out.println("stigma:\t\t"+rank_9.toString());
System.out.println("cue:\t\t"+rank_10.toString());
System.out.println("position:\t\t"+rank_11.toString());

//////////
ArrayList<ArrayList<Integer>> all_ranks= new
ArrayList<ArrayList<Integer>>();
all_ranks.add(rank_1);
all_ranks.add(rank_2);
all_ranks.add(rank_3);
all_ranks.add(rank_4);
all_ranks.add(rank_5);
all_ranks.add(rank_6);
all_ranks.add(rank_7);
all_ranks.add(rank_8);
all_ranks.add(rank_9);
all_ranks.add(rank_10);
all_ranks.add(rank_11);
ArrayList<Integer> all_rank_weights= new ArrayList<Integer>();
all_rank_weights.add(2);
all_rank_weights.add(2);
all_rank_weights.add(2);
all_rank_weights.add(2);
all_rank_weights.add(2);
all_rank_weights.add(2);
all_rank_weights.add(2);
all_rank_weights.add(2);
all_rank_weights.add(2);
all_rank_weights.add(2);
all_rank_weights.add(4);

ArrayList<Integer> final_rank=
rank_vote(all_ranks,all_rank_weights);
System.out.println("*** final_rank:\t\t"+final_rank.toString());

```

```

        int number_of_lines_in_summary=
(int) ((double)inputlines.size()*0.3);
        ArrayList<String> summary= convert_rank_string(inputlines,
number_of_lines_in_summary, final_rank);

        return summary;
    }

```

5) Evaluation Code:

```

import java.io.BufferedReader;
import java.io.DataInputStream;
import java.io.FileInputStream;
import java.io.IOException;
import java.io.InputStreamReader;
import java.util.ArrayList;

public class Evaluate {

    public static void main(String[] args) throws IOException{

        ArrayList<String>
summary_generated=readFile("C:\\Users\\Elnaz\\workspace\\Summerization\\sum1.
txt");

        ArrayList<String>
summary_ideal=readFile("C:\\Users\\Elnaz\\workspace\\Summerization\\ideall.tx
t");

        int C= find_common_sentences(summary_generated, summary_ideal);
        double Recall=(double)C/(double)summary_ideal.size();
        double Precision=(double)C/(double)summary_generated.size();

    }

    private static int find_common_sentences(ArrayList<String> sum1,
ArrayList<String> sum2) {
        System.out.println(sum1.toString());
        System.out.println("* * *");
        System.out.println(sum2.toString());
        System.out.println(sum1.size());
        System.out.println("* * *");
        System.out.println(sum2.size());

        int count=0;
        for(int i=0;i<sum1.size();i++){
            for(int j=0;j<sum2.size();j++){
                if(sum1.get(i).trim().equals(sum2.get(j).trim())){
                    count++;
                }
            }
        }
    }
}

```

```

                break;
            }
        }
    }
    System.out.print("# sentences in common: "+count);
    if(count > sum1.size() && count > sum2.size()){
        System.out.print("This is an error");
        System.exit(1);
    }
    return count;
}

public static ArrayList<String> readFile(String rulefile) throws
IOException{
    FileInputStream fstream = new FileInputStream(rulefile);
    DataInputStream in = new DataInputStream(fstream);
    BufferedReader br = new BufferedReader(new InputStreamReader(in));
    ArrayList<String> linelist=new ArrayList<String>();
    String strLine;
    while ((strLine = br.readLine()) != null) {
        linelist.add(strLine.toLowerCase());
        // System.out.println(strLine);
    }
    // System.out.println();
    return linelist;
}
}

```

6) Evaluation Results:

The following shows the result of running both Score and Rank summarization method on document2 and document 3, with the generated documents and measure precision and recall. For the document3, the result of running 2 different query sets is shown:

Document 2:

Query: Einstein star theory

Summary Generated by rank method:

scientists have put einstein's theory of gravity to a remarkable test using two distant, dead stars.

22.539915602342848

an international team studied a double pulsar system, a pair of dense neutron stars whose properties mean they act like near-perfect space clocks. 25.52284542220199

einstein's theory of general relativity (gr) describes mass and gravity as being generated by the curvature of space-time. 11.570776892068277

"we've found this double pulsar system to be an excellent laboratory for studying a wide range of physical problems," team leader professor michael kramer from the university of manchester's jodrell bank observatory, uk, told bbc news. 6.546833479901992

dead stars provide einstein test 7.693410268337176
the astronomers observed the way each stellar timepiece behaved in the other's curved space-time to run the rule over general relativity. 3.8856985275838714
the results, reported in science, match einstein's predictions to within 0.05%.

12.766051459593436

by monitoring the pulses' arrival at earth, the team was able to measure how the beams from each neutron star were being disturbed as they passed through the curved space-time near their companion. 4.054714891505896

professor kramer's team discovered the double pulsar system three years ago.

5.048797182732091

einstein's theory predicts that the double pulsar system should be emitting gravitational waves - ripples in space-time that spread out across the universe at the speed of light - as the neutron stars gradually fall in on each other. -2.982182938678478

sentences in common: 8

Recall: 0.7272727272727273

Precision: 0.8

Summary Generated by score method:

dead stars provide einstein test 25.52284542220199

scientists have put einstein's theory of gravity to a remarkable test using two distant, dead stars.

22.539915602342848

einstein's theory of general relativity (gr) describes mass and gravity as being generated by the curvature of space-time. 11.570776892068277

einstein's theory predicts that the double pulsar system should be emitting gravitational waves - ripples in space-time that spread out across the universe at the speed of light - as the neutron stars gradually fall in on each other. 6.546833479901992

an international team studied a double pulsar system, a pair of dense neutron stars whose properties mean they act like near-perfect space clocks. 7.693410268337176

the results, reported in science, match einstein's predictions to within 0.05%.

3.8856985275838714

the astronomers observed the way each stellar timepiece behaved in the other's curved space-time to run the rule over general relativity. 12.766051459593436

professor kramer's team discovered the double pulsar system three years ago.

4.054714891505896

by monitoring the pulses' arrival at earth, the team was able to measure how the beams from each neutron star were being disturbed as they passed through the curved space-time near their companion. 5.048797182732091

space ticks -2.982182938678478

sentences in common: 8

Recall: 0.72727272727273

Precision: 0.8

Summary Ideal Generated:

Dead stars provide Einstein test

Scientists have put Einstein's theory of gravity to a remarkable test using two distant, dead stars.

An international team studied a double pulsar system, a pair of dense neutron stars whose properties mean they act like near-perfect space clocks.

The astronomers observed the way each stellar timepiece behaved in the other's curved space-time to run the rule over General Relativity.

The results, reported in Science, match Einstein's predictions to within 0.05%.

Professor Kramer's team discovered the double pulsar system three years ago.

By monitoring the pulses' arrival at Earth, the team was able to measure how the beams from each neutron star were being disturbed as they passed through the curved space-time near their companion.

Einstein's theory predicts that the double pulsar system should be emitting gravitational waves - ripples in space-time that spread out across the Universe at the speed of light - as the neutron stars gradually fall in on each other.

Einstein's ideas have stood up extremely well to almost 100 years of scientific scrutiny, but researchers recognise General Relativity is likely to need at least some revision as they seek to unify its description of the large-scale Universe with quantum mechanics, the ideas that explain the realm of sub-atomic particles.

In the future, the Kramer team hopes to use its stellar clocks to directly detect gravitational waves emitted by merging supermassive black holes.

As the waves pass through the curved space-time around a network of nearby pulsars, they should leave their tell-tale imprint in the radio beam signals received on Earth.

Query: rebels italy

Document 3:

Summary Generated by rank method:

the international contact group on libya has agreed to create a temporary fund to assist rebel groups, during talks in rome. 28.03653633681808

libya: contact group creates fund for rebels 23.061677919889263

speaking in rome, italian foreign minister franco frattini said the creation of a new financial mechanism would "permit funds to be channelled effectively and transparently" to the rebels.

-0.2239463989971373

british foreign secretary william Hague insisted that any financial assistance to the rebels would not be spent on weapons. 2.6225838650947013

the group of nato countries, arab states and other nations are discussing ways to increase pressure on col muammar gaddafi. -0.9583427606533341

the rebels' transitional national council says it needs \$2bn-\$3bn (£1.2bn-£1.8bn) in the coming months for military salaries, food, medicine and other basic supplies. -0.32659829810757124

french foreign minister alain juppe says the fund should be operational within weeks.

7.522148580232469

the move followed an urgent request from the rebels, based in the eastern city of benghazi.

-7.811309721423557

sentences in common: 6

Recall: 0.46153846153846156

Precision: 0.75

Summary Generated by score method:

libya: contact group creates fund for rebels 28.03653633681808

the international contact group on libya has agreed to create a temporary fund to assist rebel groups, during talks in rome. 23.061677919889263

british foreign secretary william Hague insisted that any financial assistance to the rebels would not be spent on weapons. -0.2239463989971373

the move followed an urgent request from the rebels, based in the eastern city of benghazi.

2.6225838650947013

speaking in rome, italian foreign minister franco Frattini said the creation of a new financial mechanism would "permit funds to be channelled effectively and transparently" to the rebels.

-0.9583427606533341

the rebels' transitional national council says it needs \$2bn-\$3bn (£1.2bn-£1.8bn) in the coming months for military salaries, food, medicine and other basic supplies. -0.32659829810757124

amid scenes of panic, families were separated and the ship had to re-dock twice before finally setting sail with mostly wounded people and stranded foreign workers. 7.522148580232469

"this [money] will help them to keep basic services going... because in the east of libya they still need to be educating people, to keep public services moving and they have to meet the expenses of all that and they don't have much tax revenue at the moment," he said. -

7.811309721423557

sentences in common: 6

Recall: 0.46153846153846156

Precision: 0.75

Summary Ideal Generated:

Dead stars provide Einstein test

Scientists have put Einstein's theory of gravity to a remarkable test using two distant, dead stars.

Libya: Contact group creates fund for rebels

The international contact group on Libya has agreed to create a temporary fund to assist rebel groups, during talks in Rome.

The group of Nato countries, Arab states and other nations are discussing ways to increase pressure on Col Muammar Gaddafi.

Meanwhile, a ship has evacuated people from the besieged port of Misrata.

Speaking in Rome, Italian Foreign Minister Franco Frattini said the creation of a new financial mechanism would "permit funds to be channelled effectively and transparently" to the rebels.

The rebels' Transitional National Council says it needs \$2bn-\$3bn (£1.2bn-£1.8bn) in the coming months for military salaries, food, medicine and other basic supplies.

British Foreign Secretary William Hague insisted that any financial assistance to the rebels would not be spent on weapons.

Mr Hague said efforts would also be made to explore how Col Gaddafi's government could be prevented from exporting oil or importing refined products.

Meanwhile, US Secretary of State Hillary Clinton said the US was trying to free more than \$30bn it had frozen in Libyan assets.

Meanwhile the aid ship - Red Star - chartered by the International Organization for Migration (IOM), has managed to take 1,300 passengers to Benghazi, despite being hit by shellfire as it docked. Five people were killed.

Amid scenes of panic, families were separated and the ship had to re-dock twice before finally setting sail with mostly wounded people and stranded foreign workers.

Rebel forces in Libya hold much of the east of the country, around Benghazi, while Col Gaddafi holds most of the west.

Nato is enforcing a UN mandate to protect civilians caught in the conflict.

Query: Gaddafi italy Benghazi

Document 3:

Summary Generated by rank method:

the international contact group on libya has agreed to create a temporary fund to assist rebel groups, during talks in rome. 21.851996554564742

the group of nato countries, arab states and other nations are discussing ways to increase pressure on col muammar gaddafi. 22.909520770027562

libya: contact group creates fund for rebels -0.37610354885883623

speaking in rome, italian foreign minister franco frattini said the creation of a new financial mechanism would "permit funds to be channelled effectively and transparently" to the rebels.

8.334464453978832

the move followed an urgent request from the rebels, based in the eastern city of benghazi.

-1.110499910515033

she said the government wanted "to tap some portion of those assets owned by gaddafi and the libyan government in the united states, so we can make those funds available to help the libyan people".

-0.47875544796927016

french foreign minister alain juppe says the fund should be operational within weeks.

1.6293447574440254

meanwhile the aid ship - red star - chartered by the international organization for migration (iom), has managed to take 1,300 passengers to benghazi, despite being hit by shellfire as it docked. five people were killed.

-7.963466871285256

sentences in common: 5

Recall: 0.38461538461538464

Precision: 0.625

Summary Generated by score method:

the international contact group on libya has agreed to create a temporary fund to assist rebel groups, during talks in rome.

21.851996554564742

libya: contact group creates fund for rebels

22.909520770027562

rebel forces in libya hold much of the east of the country, around benghazi, while col gaddafi holds most of the west.

-0.37610354885883623

the group of nato countries, arab states and other nations are discussing ways to increase pressure on col muammar gaddafi.

8.334464453978832

meanwhile the aid ship - red star - chartered by the international organization for migration (iom), has managed to take 1,300 passengers to benghazi, despite being hit by shellfire as it docked. five people were killed.

-1.110499910515033

the move followed an urgent request from the rebels, based in the eastern city of benghazi.

-0.47875544796927016

amid scenes of panic, families were separated and the ship had to re-dock twice before finally setting sail with mostly wounded people and stranded foreign workers. "this [money] will help them to keep basic services going... because in the east of libya they still need to be educating people, to keep public services moving and they have to meet the expenses of all that and they don't have much tax revenue at the moment," he said.

-7.963466871285256

sentences in common: 6

Recall: 0.46153846153846156

Precision: 0.75

Summary Ideal Generated:

Dead stars provide Einstein test

Scientists have put Einstein's theory of gravity to a remarkable test using two distant, dead stars.

Libya: Contact group creates fund for rebels

The international contact group on Libya has agreed to create a temporary fund to assist rebel groups, during talks in Rome.

The group of Nato countries, Arab states and other nations are discussing ways to increase pressure on Col Muammar Gaddafi.

Meanwhile, a ship has evacuated people from the besieged port of Misrata.

Speaking in Rome, Italian Foreign Minister Franco Frattini said the creation of a new financial mechanism would "permit funds to be channelled effectively and transparently" to the rebels.

The rebels' Transitional National Council says it needs \$2bn-\$3bn (£1.2bn-£1.8bn) in the coming months for military salaries, food, medicine and other basic supplies.

British Foreign Secretary William Hague insisted that any financial assistance to the rebels would not be spent on weapons.

Mr Hague said efforts would also be made to explore how Col Gaddafi's government could be prevented from exporting oil or importing refined products.

Meanwhile, US Secretary of State Hillary Clinton said the US was trying to free more than \$30bn it had frozen in Libyan assets.

Meanwhile the aid ship - Red Star - chartered by the International Organization for Migration (IOM), has managed to take 1,300 passengers to Benghazi, despite being hit by shellfire as it docked. Five people were killed.

Amid scenes of panic, families were separated and the ship had to re-dock twice before finally setting sail with mostly wounded people and stranded foreign workers.

Rebel forces in Libya hold much of the east of the country, around Benghazi, while Col Gaddafi holds most of the west.

Nato is enforcing a UN mandate to protect civilians caught in the conflict.
