Filename	Size	Description of the Textfile
sre01.txt	11278	SRE: The Saga Of The Best SRE Game Ever Played! By Josh Renaud
sre02.txt	5862	Solar Realms Elite: The True Story of the Unsung Heroes, by Josh Renaud
sre03.txt	8555	Solar Realms Elite: Ultra's Untold Story by Josh Renaud
sre04.txt	44198	Solar Realms Elite IV: The Confrontation, by Josh Renaud
sre05.txt	20787	Solar Realms Elite V: The Underground, by Josh Renaud
sre06.txt	26731	Solar Realms Elite VI: The Alliance Restored, by Josh Renaud
sre07.txt	23597	Solar Realms Elite 7: Petros, by Josh Renaud
sre08.txt	33170	Solar Realms Elite VIII: Kazik, by Josh Renaud
sre09.txt	26073	Solar Realms Elite IX: Survival of the Fittest, by Josh Renaud
sre10.txt	25725	Solar Realms Elite X: Legacies, by Josh Renaud
sre_feqh.txt	20054	Solar Realms Elite: The Feqh Galaxy, by Josh Renaud
sre_finl.txt	33158	Solar Realms Elite: The Finale by Josh Renaud
sre_sei.txt	20753	Solar Realms Elite: Galaxy Sei, by Josh Renaud
sretrade.txt	9008	The SRE Commerce and Trade Theories, by Josh Renaud
srex.txt	37128	Solar Realms Elite: X1 and X2, by Josh Renaud

There are 15 files for

```
<TR VALIGN=TOP><TD ALIGN=TOP><A HREF="13chil.txt">13chil.txt</A>  <tab t
o=T><TD> 8457<BR><TD> The Story of the Sly Fox
<TR VALIGN=TOP><TD ALIGN=TOP><A HREF="14.lws">14.lws</A>  <tab to=T><TD>
5261<BR><TD> A Smart Bomb with a Language Parser
<TR VALIGN=TOP><TD ALIGN=TOP><A HREF="16.lws">16.lws</A>  <tab to=T><TD>
15294<BR><TD> Two Guys in a Garage, by M. Pshota
<TR VALIGN=TOP><TD ALIGN=TOP><A HREF="17.lws">17.lws</A>  <tab to=T><TD>
10853<BR><TD> The Early Days of a High-Tech Start-up are Magic (November
18, 1991) by M. Peshota
<TR VALIGN=TOP><TD ALIGN=TOP><A HREF="18.lws">18.lws</A>  <tab to=T><TD>
26624<BR><TD> The Couch, the File Cabinet, and the Calendar, by M. Pesho
ta (December 9, 1991)
<TR VALIGN=TOP><TD ALIGN=TOP><A HREF="19.lws">19.lws</A>  <tab to=T><TD>
17902<BR><TD> Engineering the Future of American Technology by M. Peshot
a (January 5, 1992)
<TR VALIGN=TOP><TD ALIGN=TOP><A HREF="20.lws">20.lws</A>  <tab to=T><TD>
13588<BR><TD> What Research and Development Was Always Meant to Be, by
M. Peshota
```

```
1   print(stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you',
'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'sh
t's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves
t', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'k
ng', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'i
f', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into',
e', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'c
e', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'bc
me', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than',
'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 'r
n', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't
"isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn', "ne
n't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn',
```

```
1   num2words(100500)
```

```
'one hundred thousand, five hundred'
```

```
def preprocess(data):
    data = convert_lower_case(data)
    data = remove_punctuation(data)
    data = remove_apostrophe(data)
    data = remove_single_characters(data)
    data = convert_numbers(data)
    data = remove_stop_words(data)
    data = stemming(data)
    data = remove_punctuation(data)
    data = convert_numbers(data)
```

```
1  for i in DF:
2      DF[i] = len(DF[i])
3  DF
```

```
{'sharewar': 1,
 'trial': 1,
 'project': 4,
 'freewar': 1,
 'need': 6,
 'support': 2,
 'continu': 4,
 'one': 10,
 'hundr': 8,
```

```
1  total_vocab = [x for x in DF]
2  print(total_vocab)
```

```
['sharewar', 'trial', 'project', 'freewar', 'ne
 'north', 'jim', 'prentic', 'copyright', 'thousa
 c', 'phrase', 'spoken', 'mumbl', 'thought', 'in
 'map', 'label', 'degr', 'presenc', 'indic', 'hi
 ct', 'mind', 'intern', 'border', 'writer', 'poe
 le', 'man', 'eat', 'mosquito', 'murder', 'hord
 'dog', 'stori', 'record', 'break', 'trout', 'wa
 'forest', 'crash', 'moo', 'tear', 'brush', 'tre
 uck', 'feed', 'quiet', 'pond', 'placid', 'bay',
 r', 'authent', 'northern', 'scene', 'wildlif',
 d', 'person', 'live', 'farther', 'becom', 'appa
```

Matching Score

Query: Without the drive of Rebeccah's insistence, Kate lost her momentum. She stood next a slatted oak bench, canisters still clutched, surveying

['without', 'drive', 'rebeccah', 'insist', 'kate', 'lost', 'momentum', 'stood', 'next', 'slat', 'oak', 'bench', 'canist', 'still', 'clutch', 'survey']

[166, 200, 352, 433, 211, 350, 175, 187, 188, 294]

Cosine Similarity

Query: Without the drive of Rebeccah's insistence, Kate lost her momentum. She stood next a slatted oak bench, canisters still clutched, surveying

['without', 'drive', 'rebeccah', 'insist', 'kate', 'lost', 'momentum', 'stood', 'next', 'slat', 'oak', 'bench', 'canist', 'still', 'clutch', 'survey']

[200 166 433 175 169 402 211 87 151 369]

Long Query

Matching Score

Query: And then she recalled his stiff body stretched out in the little bed over the garages. Another pearl had come
loose from the strand, seeming to want to search out its old home in a far away oyster bed. She would have those pear
ls laid out n

['recal', 'stiff', 'bodi', 'stretch', 'littl', 'bed', 'garag', 'anoth', 'pearl', 'come', 'loo', 'strand', 'seem', 'wa
nt', 'search', 'old', 'home', 'far', 'away', 'oyster', 'bed', 'would', 'pearl', 'laid']

[352, 351, 43, 269, 17, 272, 9, 67, 416, 208]

Cosine Similarity

Query: And then she recalled his stiff body stretched out in the little bed over the garages. Another pearl had come
loose from the strand, seeming to want to search out its old home in a far away oyster bed. She would have those pear
ls laid out n

['recal', 'stiff', 'bodi', 'stretch', 'littl', 'bed', 'garag', 'anoth', 'pearl', 'come', 'loo', 'strand', 'seem', 'wa
nt', 'search', 'old', 'home', 'far', 'away', 'oyster', 'bed', 'would', 'pearl', 'laid']

[16   3 200 151 65 364 169 87 27 82]