# High-level description of Code

1. **Stack class:**

# Define a stack data structure with a fixed maximum size

Class Stack:

Initialise STACK\_SIZE as a constant

Initialise stack as an empty list

Initialise top as -1 # Indicates an empty stack

# Method to add an element to the stack

Method push(word)

If top < STACK\_SIZE - 1 then

Append word to stack

Increment top

Else

print "Stack is full."

end if

# Method to remove and return the top element

Method pop()

If top >= 0 then

Remove and return stack[top]

Decrement top

Else

print "Stack is empty."

end if

# Check if the stack is empty

Method is\_empty()

return top == -1

1. **Define Comparator Functions**

# Comparator function to compare word-count pairs first by frequency, then alphabetically

Method Compare(item1, item2)

# Check if the frequencies are different

If item1 frequency is not equal to item2 frequency, then

Return item1[1] – item2[1]

Else

# If frequencies are the same, compare the strings

Return (item2[0] > item1[0]) – (item2[0] < item1[0]) # Compare words alphabetically

# Comparator function for alphabetical comparison of words

Method alphabetical\_compare(item1, item2)

If item1 < item2 then # Compare words alphabetically

Return -1 # item1 comes first alphabetically

Else If item1 > item2 then

Return 1 # item1 comes later alphabetically

Else

Return 0 # Both words are identical

End if

1. **Heap Functions**

# Function to sift down an element in the heap to maintain the heap property

Method siftdown(arr, start, end, compare)

Set root to start # Begin with the root of the loop

Repeat

Set child to 2 \* root + 1 # Calculate left child index

If child > end, then

Stop # If child is out of bounds, stop

# Check if right child exists and is smaller

If child + 1 <= end and compare(arr[child], arr[child + 1]) > 0 then

Increment child # Point to the smaller child

If compare(arr[root], arr[child]) > 0 then # If root is greater than the smaller child

Swap arr[root] and arr[child]

Set root to child # Move down to the child node

Else

Stop

End If

Forever

# Function to build a heap from an unsorted array

Function makeheap(ar, compare)

Set length to length of arr # Get the number of elements in the array

Set start to (length – 2) // 2 # Start with the last parent node

While start >= 0

Call siftdown(arr, start, length – 1, compare)

Decrement start # Move to the previous parent node

End while

# Function to perform heapsort on an array

Function heapsort(arr, compare)

Call makeheap(arr, compare)

Set end to the length of arr – 1

While end > 0 do

Swap arr[end] and arr[0] # Move the largest element to the end

Decrement end

Call siftdown(arr, 0, end, compare) # Restore the heap property

End while

1. **Process File Function**

Function process\_file()

Prompt user for filename

Initialise word\_list and count\_list as empty lists # 2 lists to store the words and their counts respectively

Initialise ascii\_letters as a list of lowercase ASCII characters # Literal containing all lowercase alphabetic characters

Try, then

# Reading file line by line.

Open file with filename in read mode

For each line in the file, do

Convert the line to lowercase

Split line in words

For each word in the line

Clean word to contain only alphabetic characters

If the word is not empty, then

If the word is in word\_list, then

Increment corresponding count in count\_list

Else

Append word to word\_list and 1 to count\_list

End if

End If

Catch FileNotFound Exception

Print “Error: The file is not found.”

Catch UnicodeDecode Exception

Print “Error: The file is not UTF-8 encoded.”

# Create a list of word-count pairs and sort then

Create sorted\_words as a list of (word, count) pairs from word\_list and count\_list

Call heapsort(sorted\_words, compare)

# Display the top 10 words by frequency

Print “First 10 words by frequency:”

For the first 10 items in sorted\_words, do

Print word and count

End for

# Display the last ten words by frequency

Print “Last ten words by frequency:”  
For the last 10 items in sorted\_words, do

Print word and count

End for

# Retrieve all unique words

Create unique\_words\_list as a list of words with a count of 1

# Sort unique words using heapsort (alphabetically)

Call heapsort(unique\_words\_list, alphabetical\_compare)

Create unique\_words as Stack.

For each word in unique\_words\_list, do

Call unique\_words.push(word)

Print “Unique words sorted alphabetically:”

While not unique\_words is not empty do

Print unique\_words.pop()

# Complexity Analysis with Big-O Notation

## Heap Operations:

* **siftdown(arr, start, end, m\_compare):** O(log n) - Each siftdown operation can propagate down the height of the heap (logarithmic height).
* **makeheap(arr, m\_compare):** O(n) - Building a heap from an unsorted array takes linear time.
* **heapsort(arr, m\_compare):** O(n log n) - Heapsort involves building the heap (O(n)) and performing n siftdown operations O(n log n) .

# Data Structures Used

1. **Dictionary (`words\_count\_pairs`):**

**Reason:** To store word-count pairs efficiently. Dictionaries are ideal for this purpose as they provide fast average-time complexity for insertions and lookups O(1). This allows for quick accumulation of word counts, which can later be sorted as needed.

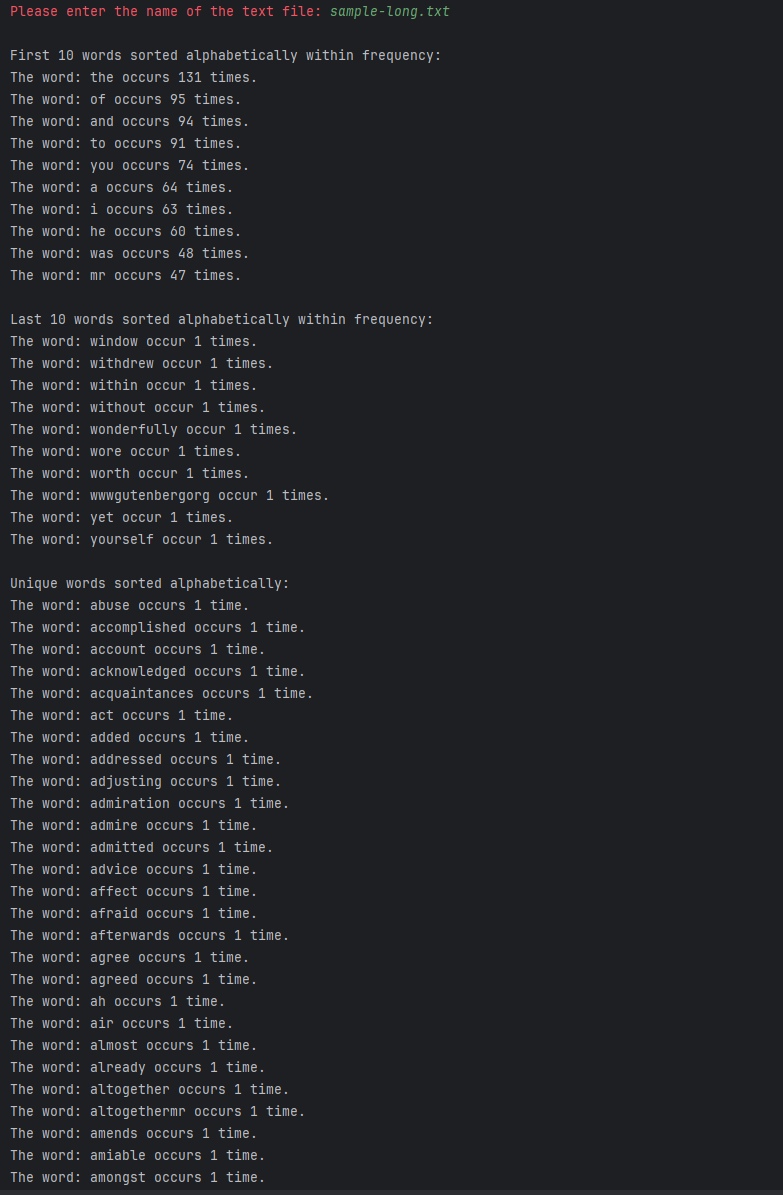
1. **Stack (`stack` class):**

**Reason:** Manage unique words in a LIFO (Last-In-First-Out) manner. The stack stores and retrieves unique words in reverse order, providing an easy way to print them in sorted order after heapsort. This approach is particularly useful when you need to perform operations on words before sorting them using heapsort, ensuring they can be easily printed in the desired order.

1. **Heap (Implicit in the Heapsort Functions):**

**Reason:** Sort the word-count pairs and unique words efficiently. Heaps are used because heapsort provides an efficient O(n log n) sorting algorithm suitable for this context. Using a heap ensures that sorting is handled effectively and efficiently, even for large datasets.

## Execution snapshot

-

## Output

C:\Users\Michael\anaconda3\envs\CSCI203\python.exe "C:\Users\Michael\OneDrive - University of Wollongong\UOW\Bachelor of Computer Science\Year2\Spring\CSCI203 - Algorithms and Data Structures\Assessments\Assignment 1\main.py"

Please enter the name of the text file: sample-long.txt

First 10 words sorted alphabetically within frequency:

The word: the occurs 131 times.

The word: of occurs 95 times.

The word: and occurs 94 times.

The word: to occurs 91 times.

The word: you occurs 74 times.

The word: a occurs 64 times.

The word: i occurs 63 times.

The word: he occurs 60 times.

The word: was occurs 48 times.

The word: mr occurs 47 times.

Last 10 words sorted alphabetically within frequency:

The word: window occurs 1 times.

The word: withdrew occurs 1 times.

The word: within occurs 1 times.

The word: without occurs 1 times.

The word: wonderfully occurs 1 times.

The word: wore occurs 1 times.

The word: worth occurs 1 times.

The word: wwwgutenbergorg occurs 1 times.

The word: yet occurs 1 times.

The word: yourself occurs 1 times.

Unique words sorted alphabetically:

The word: abuse occurs 1 time.

The word: accomplished occurs 1 time.

The word: account occurs 1 time.

The word: acknowledged occurs 1 time.

The word: acquaintances occurs 1 time.

The word: act occurs 1 time.

The word: added occurs 1 time.

The word: addressed occurs 1 time.

The word: adjusting occurs 1 time.

The word: admiration occurs 1 time.

The word: admire occurs 1 time.

The word: admitted occurs 1 time.

The word: advice occurs 1 time.

The word: affect occurs 1 time.

The word: afraid occurs 1 time.

The word: afterwards occurs 1 time.

The word: agree occurs 1 time.

The word: agreed occurs 1 time.

The word: ah occurs 1 time.

The word: air occurs 1 time.

The word: almost occurs 1 time.

The word: already occurs 1 time.

The word: altogether occurs 1 time.

The word: altogethermr occurs 1 time.

The word: amends occurs 1 time.

The word: amiable occurs 1 time.

The word: amongst occurs 1 time.

The word: amusement occurs 1 time.

The word: angry occurs 1 time.

The word: ankle occurs 1 time.

The word: anyone occurs 1 time.

The word: anywhere occurs 1 time.

The word: arrival occurs 1 time.

The word: arrived occurs 1 time.

The word: ascertaining occurs 1 time.

The word: assemblies occurs 1 time.

The word: assistance occurs 1 time.

The word: assuring occurs 1 time.

The word: astonishment occurs 1 time.

The word: attacked occurs 1 time.

The word: attention occurs 1 time.

The word: author occurs 1 time.

The word: away occurs 1 time.

The word: aye occurs 1 time.

The word: back occurs 1 time.

The word: barefaced occurs 1 time.

The word: behaviour occurs 1 time.

The word: beheld occurs 1 time.

The word: behind occurs 1 time.

The word: believe occurs 1 time.

The word: best occurs 1 time.

The word: bingleys occurs 1 time.

The word: bit occurs 1 time.

The word: bitterness occurs 1 time.

The word: black occurs 1 time.

The word: blue occurs 1 time.

The word: book occurs 1 time.

The word: books occurs 1 time.

The word: boulanger occurs 1 time.

The word: branch occurs 1 time.

The word: bring occurs 1 time.

The word: brotherinlaw occurs 1 time.

The word: brought occurs 1 time.

The word: called occurs 1 time.

The word: caprice occurs 1 time.

The word: care occurs 1 time.

The word: cases occurs 1 time.

The word: catching occurs 1 time.

The word: catherine occurs 1 time.

The word: certain occurs 1 time.

The word: chaise occurs 1 time.

The word: chance occurs 1 time.

The word: charming occurs 1 time.

The word: check occurs 1 time.

The word: children occurs 1 time.

The word: choose occurs 1 time.

The word: chooses occurs 1 time.

The word: circulation occurs 1 time.

The word: circumspection occurs 1 time.

The word: closed occurs 1 time.

The word: coat occurs 1 time.

The word: coldly occurs 1 time.

The word: comforted occurs 1 time.

The word: company occurs 1 time.

The word: compared occurs 1 time.

The word: conceited occurs 1 time.

The word: conjecturing occurs 1 time.

The word: consent occurs 1 time.

The word: consequence occurs 1 time.

The word: consequently occurs 1 time.

The word: consideration occurs 1 time.

The word: considered occurs 1 time.

The word: consisted occurs 1 time.

The word: contain occurs 1 time.

The word: continued occurs 1 time.

The word: contrast occurs 1 time.

The word: conversation occurs 1 time.

The word: copy occurs 1 time.

The word: cordial occurs 1 time.

The word: cost occurs 1 time.

The word: coughing occurs 1 time.

The word: coughs occurs 1 time.

The word: country occurs 1 time.

The word: courses occurs 1 time.

The word: cousin occurs 1 time.

The word: credit occurs 1 time.

The word: crown occurs 1 time.

The word: curiosity occurs 1 time.

The word: dances occurs 1 time.

The word: days occurs 1 time.

The word: deal occurs 1 time.

The word: declare occurs 1 time.

The word: declared occurs 1 time.

The word: decline occurs 1 time.

The word: declined occurs 1 time.

The word: deep occurs 1 time.

The word: deferred occurs 1 time.

The word: deigned occurs 1 time.

The word: delight occurs 1 time.

The word: derbyshire occurs 1 time.

The word: desire occurs 1 time.

The word: determined occurs 1 time.

The word: determining occurs 1 time.

The word: develop occurs 1 time.

The word: different occurs 1 time.

The word: difficult occurs 1 time.

The word: disappointed occurs 1 time.

The word: disclosed occurs 1 time.

The word: disconcerted occurs 1 time.

The word: discontented occurs 1 time.

The word: discovered occurs 1 time.

The word: discretion occurs 1 time.

The word: disgust occurs 1 time.

The word: dislike occurs 1 time.

The word: dispatched occurs 1 time.

The word: disposition occurs 1 time.

The word: distant occurs 1 time.

The word: distinguished occurs 1 time.

The word: dont occurs 1 time.

The word: door occurs 1 time.

The word: draw occurs 1 time.

The word: dresses occurs 1 time.

The word: drew occurs 1 time.

The word: during occurs 1 time.

The word: earliest occurs 1 time.

The word: early occurs 1 time.

The word: easy occurs 1 time.

The word: either occurs 1 time.

The word: elegant occurs 1 time.

The word: else occurs 1 time.

The word: eluded occurs 1 time.

The word: emphatic occurs 1 time.

The word: employed occurs 1 time.

The word: enduring occurs 1 time.

The word: engage occurs 1 time.

The word: engaged occurs 1 time.

The word: england occurs 1 time.

The word: enjoy occurs 1 time.

The word: enquired occurs 1 time.

The word: entering occurs 1 time.

The word: entrance occurs 1 time.

The word: equally occurs 1 time.

The word: escape occurs 1 time.

The word: establishment occurs 1 time.

The word: estate occurs 1 time.

The word: etc occurs 1 time.

The word: event occurs 1 time.

The word: exaggeration occurs 1 time.

The word: excessively occurs 1 time.

The word: exclamation occurs 1 time.

The word: expectations occurs 1 time.

The word: expected occurs 1 time.

The word: experience occurs 1 time.

The word: extracts occurs 1 time.

The word: extraordinary occurs 1 time.

The word: extremely occurs 1 time.

The word: eye occurs 1 time.

The word: fall occurs 1 time.

The word: falling occurs 1 time.

The word: families occurs 1 time.

The word: family occurs 1 time.

The word: fancied occurs 1 time.

The word: fancy occurs 1 time.

The word: fancying occurs 1 time.

The word: fashion occurs 1 time.

The word: fastidious occurs 1 time.

The word: fatigued occurs 1 time.

The word: favourable occurs 1 time.

The word: fear occurs 1 time.

The word: fears occurs 1 time.

The word: features occurs 1 time.

The word: felt occurs 1 time.

The word: fifth occurs 1 time.

The word: figure occurs 1 time.

The word: find occurs 1 time.

The word: finery occurs 1 time.

The word: fixed occurs 1 time.

The word: flatter occurs 1 time.

The word: flying occurs 1 time.

The word: fond occurs 1 time.

The word: forbidding occurs 1 time.

The word: forget occurs 1 time.

The word: forms occurs 1 time.

The word: fortnights occurs 1 time.

The word: fourth occurs 1 time.

The word: fretfully occurs 1 time.

The word: gave occurs 1 time.

The word: gentleman occurs 1 time.

The word: gentlemanlike occurs 1 time.

The word: given occurs 1 time.

The word: gods occurs 1 time.

The word: going occurs 1 time.

The word: goodhumoured occurs 1 time.

The word: goodlooking occurs 1 time.

The word: got occurs 1 time.

The word: gown occurs 1 time.

The word: gratified occurs 1 time.

The word: grieved occurs 1 time.

The word: grownup occurs 1 time.

The word: handsomer occurs 1 time.

The word: happily occurs 1 time.

The word: hat occurs 1 time.

The word: hate occurs 1 time.

The word: heart occurs 1 time.

The word: hearty occurs 1 time.

The word: heavens occurs 1 time.

The word: hertfordshire occurs 1 time.

The word: highly occurs 1 time.

The word: horrid occurs 1 time.

The word: horse occurs 1 time.

The word: house occurs 1 time.

The word: housekeeping occurs 1 time.

The word: hursts occurs 1 time.

The word: hypocritical occurs 1 time.

The word: idea occurs 1 time.

The word: ideas occurs 1 time.

The word: ignorant occurs 1 time.

The word: ill occurs 1 time.

The word: im occurs 1 time.

The word: imagine occurs 1 time.

The word: immediately occurs 1 time.

The word: included occurs 1 time.

The word: information occurs 1 time.

The word: ingenious occurs 1 time.

The word: inhabitants occurs 1 time.

The word: instead occurs 1 time.

The word: insufficient occurs 1 time.

The word: insupportable occurs 1 time.

The word: intelligence occurs 1 time.

The word: intended occurs 1 time.

The word: interrupted occurs 1 time.

The word: introduction occurs 1 time.

The word: its occurs 1 time.

The word: janes occurs 1 time.

The word: join occurs 1 time.

The word: joke occurs 1 time.

The word: joy occurs 1 time.

The word: keep occurs 1 time.

The word: king occurs 1 time.

The word: kingdom occurs 1 time.

The word: knowledge occurs 1 time.

The word: lace occurs 1 time.

The word: laid occurs 1 time.

The word: laws occurs 1 time.

The word: learnt occurs 1 time.

The word: least occurs 1 time.

The word: left occurs 1 time.

The word: less occurs 1 time.

The word: library occurs 1 time.

The word: license occurs 1 time.

The word: likely occurs 1 time.

The word: likes occurs 1 time.

The word: lines occurs 1 time.

The word: live occurs 1 time.

The word: lived occurs 1 time.

The word: london occurs 1 time.

The word: londonhis occurs 1 time.

The word: longbourn occurs 1 time.

The word: looking occurs 1 time.

The word: lose occurs 1 time.

The word: loved occurs 1 time.

The word: making occurs 1 time.

The word: mamma occurs 1 time.

The word: maria occurs 1 time.

The word: matter occurs 1 time.

The word: meaning occurs 1 time.

The word: meant occurs 1 time.

The word: meet occurs 1 time.

The word: mention occurs 1 time.

The word: mentioned occurs 1 time.

The word: met occurs 1 time.

The word: michaelmas occurs 1 time.

The word: mien occurs 1 time.

The word: might occurs 1 time.

The word: mind occurs 1 time.

The word: minds occurs 1 time.

The word: mistake occurs 1 time.

The word: mixture occurs 1 time.

The word: moment occurs 1 time.

The word: monday occurs 1 time.

The word: morris occurs 1 time.

The word: name occurs 1 time.

The word: near occurs 1 time.

The word: neglect occurs 1 time.

The word: neighbour occurs 1 time.

The word: nervous occurs 1 time.

The word: new occurs 1 time.

The word: newcomers occurs 1 time.

The word: news occurs 1 time.

The word: noble occurs 1 time.

The word: nobody occurs 1 time.

The word: none occurs 1 time.

The word: nor occurs 1 time.

The word: north occurs 1 time.

The word: number occurs 1 time.

The word: objection occurs 1 time.

The word: observing occurs 1 time.

The word: occasionally occurs 1 time.

The word: odd occurs 1 time.

The word: office occurs 1 time.

The word: often occurs 1 time.

The word: old occurs 1 time.

The word: online occurs 1 time.

The word: opinion occurs 1 time.

The word: out occurs 1 time.

The word: overscrupulous occurs 1 time.

The word: park occurs 1 time.

The word: part occurs 1 time.

The word: particular occurs 1 time.

The word: particularly occurs 1 time.

The word: passed occurs 1 time.

The word: people occurs 1 time.

The word: person occurs 1 time.

The word: persuade occurs 1 time.

The word: pieces occurs 1 time.

The word: planned occurs 1 time.

The word: playful occurs 1 time.

The word: pleasantly occurs 1 time.

The word: pleasing occurs 1 time.

The word: pleasure occurs 1 time.

The word: poor occurs 1 time.

The word: popularity occurs 1 time.

The word: preference occurs 1 time.

The word: press occurs 1 time.

The word: pretend occurs 1 time.

The word: promised occurs 1 time.

The word: pronounced occurs 1 time.

The word: property occurs 1 time.

The word: protested occurs 1 time.

The word: proud occurs 1 time.

The word: proudest occurs 1 time.

The word: punishment occurs 1 time.

The word: qualities occurs 1 time.

The word: questions occurs 1 time.

The word: quick occurs 1 time.

The word: quickness occurs 1 time.

The word: quieted occurs 1 time.

The word: quieter occurs 1 time.

The word: raised occurs 1 time.

The word: raptures occurs 1 time.

The word: rather occurs 1 time.

The word: read occurs 1 time.

The word: really occurs 1 time.

The word: recommend occurs 1 time.

The word: reflection occurs 1 time.

The word: regardless occurs 1 time.

The word: related occurs 1 time.

The word: remained occurs 1 time.

The word: reply occurs 1 time.

The word: resentfully occurs 1 time.

The word: resentment occurs 1 time.

The word: reserve occurs 1 time.

The word: respect occurs 1 time.

The word: restrictions occurs 1 time.

The word: reuse occurs 1 time.

The word: ridiculous occurs 1 time.

The word: rightful occurs 1 time.

The word: rode occurs 1 time.

The word: round occurs 1 time.

The word: rudeness occurs 1 time.

The word: sakes occurs 1 time.

The word: sarcastic occurs 1 time.

The word: sat occurs 1 time.

The word: satisfactory occurs 1 time.

The word: save occurs 1 time.

The word: says occurs 1 time.

The word: scarcity occurs 1 time.

The word: scolding occurs 1 time.

The word: secondhand occurs 1 time.

The word: seek occurs 1 time.

The word: seemed occurs 1 time.

The word: seen occurs 1 time.

The word: selfish occurs 1 time.

The word: sensible occurs 1 time.

The word: servants occurs 1 time.

The word: serving occurs 1 time.

The word: setdowns occurs 1 time.

The word: settling occurs 1 time.

The word: seven occurs 1 time.

The word: several occurs 1 time.

The word: share occurs 1 time.

The word: sharpened occurs 1 time.

The word: shocking occurs 1 time.

The word: shut occurs 1 time.

The word: sick occurs 1 time.

The word: sight occurs 1 time.

The word: silly occurs 1 time.

The word: sit occurs 1 time.

The word: sitting occurs 1 time.

The word: six occurs 1 time.

The word: sixth occurs 1 time.

The word: skill occurs 1 time.

The word: smiles occurs 1 time.

The word: solace occurs 1 time.

The word: somebody occurs 1 time.

The word: somewhat occurs 1 time.

The word: sorry occurs 1 time.

The word: speak occurs 1 time.

The word: speaking occurs 1 time.

The word: spirits occurs 1 time.

The word: splendid occurs 1 time.

The word: spoke occurs 1 time.

The word: sprained occurs 1 time.

The word: stared occurs 1 time.

The word: starting occurs 1 time.

The word: step occurs 1 time.

The word: stoutly occurs 1 time.

The word: stranger occurs 1 time.

The word: stress occurs 1 time.

The word: struck occurs 1 time.

The word: stupid occurs 1 time.

The word: suddenly occurs 1 time.

The word: suffer occurs 1 time.

The word: sufficient occurs 1 time.

The word: suiting occurs 1 time.

The word: suppositions occurs 1 time.

The word: surely occurs 1 time.

The word: surmises occurs 1 time.

The word: surpassing occurs 1 time.

The word: surrounding occurs 1 time.

The word: talk occurs 1 time.

The word: talked occurs 1 time.

The word: tall occurs 1 time.

The word: tallest occurs 1 time.

The word: tear occurs 1 time.

The word: teasing occurs 1 time.

The word: temper occurs 1 time.

The word: tempt occurs 1 time.

The word: terms occurs 1 time.

The word: these occurs 1 time.

The word: third occurs 1 time.

The word: those occurs 1 time.

The word: thought occurs 1 time.

The word: threeandtwenty occurs 1 time.

The word: throw occurs 1 time.

The word: tide occurs 1 time.

The word: times occurs 1 time.

The word: tiresome occurs 1 time.

The word: title occurs 1 time.

The word: tolerable occurs 1 time.

The word: tomorrow occurs 1 time.

The word: toward occurs 1 time.

The word: towards occurs 1 time.

The word: trimming occurs 1 time.

The word: tumult occurs 1 time.

The word: turned occurs 1 time.

The word: turning occurs 1 time.

The word: unaffected occurs 1 time.

The word: uncertain occurs 1 time.

The word: uncommonly occurs 1 time.

The word: under occurs 1 time.

The word: understand occurs 1 time.

The word: understanding occurs 1 time.

The word: universally occurs 1 time.

The word: unless occurs 1 time.

The word: unlucky occurs 1 time.

The word: unreserved occurs 1 time.

The word: unworthy occurs 1 time.

The word: upper occurs 1 time.

The word: using occurs 1 time.

The word: various occurs 1 time.

The word: venture occurs 1 time.

The word: vexed occurs 1 time.

The word: vexing occurs 1 time.

The word: village occurs 1 time.

The word: violent occurs 1 time.

The word: visiting occurs 1 time.

The word: waited occurs 1 time.

The word: walking occurs 1 time.

The word: wasting occurs 1 time.

The word: wayswith occurs 1 time.

The word: week occurs 1 time.

The word: whatsoever occurs 1 time.

The word: whichever occurs 1 time.

The word: whom occurs 1 time.

The word: wifes occurs 1 time.

The word: window occurs 1 time.

The word: withdrew occurs 1 time.

The word: within occurs 1 time.

The word: without occurs 1 time.

The word: wonderfully occurs 1 time.

The word: wore occurs 1 time.

The word: worth occurs 1 time.

The word: wwwgutenbergorg occurs 1 time.

The word: yet occurs 1 time.

The word: yourself occurs 1 time.

Process finished with exit code 0