

Please extensively and carefully test your assignment.

In your write up, please discuss these issues: What are the boundary cases?

The boundary cases include cases like the following:

File name not given

File is not readable like jpg, tiff etc.,

File format is not as given in sample file: title,http://URL

Eg: Title is empty, URL is empty, empty lines in between the text, ",http://" missing,

URL not functional

Title contains commas and special characters

More than 2 values in one line in the text file.

Less than 2 values in one line in the text file

Space or tabs instead of title or URL

Word does not occur in any books

I have used conditions and exceptions to handle such boundary cases. Like checking for empty titles, empty URL, length of the values read, if an empty line was read. In those cases, the value which was read was not added to the dictionary and count of the total books was not increased.

I also used a set of default books and titles, when none of the entries in the catalog file was found to be valid.

In addition to this, once the valid values are entered in the dictionary of books, each URL is opened and read for the contents. If the URL is not valid or if there was an error reading the URL or if the URL was empty, then a message is displayed that the URL is not valid, but the title and URL is still included in the catalog and titles list.

If none of the URL read had contents, then default values were used, to avoid showing no entries were found to the user. This default contents can be seen using the <catalog> and <titles> commands.

What happens if catalog.txt comes in an unexpected format?

If catalog.txt comes in an unexpected format, then the value corresponding to the unexpected format is ignored and not added to the dictionary of books, titles or catalog. If none of the entries in the catalog.txt file is valid, then default values are used, otherwise no content will be present.

What happens if a link is broken or the contents come in an unexpected format? etc.

If the link is broken or contents could not be read, an exception was added to catch that exception and message displayed to the user that the URL was not valid. The title and the catalog still has this file but its contents are empty. So, all the words in the dictionary will have 0 corresponding to this book title.

Test strategies used:

The program was tested using the following conditions

Assumption:

The format of the content should be of the form title,http://URL – anything else will be ignored and not added to the books.

1. If the file name is not given – Use default values.
2. If the file is not readable – Use default values.
3. If the file is not in the correct format – Use default values
4. If the URL is not active or not reachable – add it to the catalog as empty file.
5. If the syntax has more than 2 items being split on ",http://" in a line except (name and url) – that value will be ignored.

"I have the split on ,http://". So if there are other items without this format this would go to the title (and stay as such) or go to the URL (resulting in failed URL read with empty contents)

Eg: Three Felines and One Canine

,http://people.ischool.berkeley.edu/~tygar/for.i206/3cats1dog.txt,http://hello.txt

Here 3 values are present after split – so this entry is ignored.

Eg: Three Felines and One Canine,

www.ebooks.com,http://people.ischool.berkeley.edu/~tygar/for.i206/3cats1dog.txt,http://hello.txt

Here 2 values on split – title is "Three Felines and One Canine, www.ebooks.com" URL is

<http://people.ischool.berkeley.edu/~tygar/for.i206/3cats1dog.txt>,<http://hello.txt>

Eg: Three, Felines, and One Canine <http://people.ischool.berkeley.edu/~tygar/for.i206/3cats1dog.txt>

Here only 1 value after split – this value is ignored.

6. If the title is empty “ , <http://people.ischool.berkeley.edu/~tygar/for.i206/3cats1dog.txt>” - value is ignored as the title is empty.
7. If the URL is empty “Three Felines and One Canine,” - value is ignored as the URL is empty.
8. If there is a blank line in between – the check for length of the values read will catch this.
9. Titles with commas – split is on “,http://” – so title is intact
Eg: Three, Felines, and One Canine ,http://people.ischool.berkeley.edu/~tygar/for.i206/3cats1dog.txt
10. If all the URL given in the given text file is not readable and no contents were found for all the URL, then again default values are used.
Eg:
Three, Felines, and One Canine,http://people.ischool.berkeley.edu/~tygar/for.i206/3cats1dog.txt,http://hello.txt
Two Felines and Two Canines,http://people.ischool.berkeley.edu/~tygar/for.i206/2cats2dogs.txt,123
One Feline and Three Canines and One
Turtle,http://people.ischool.berkeley.edu/~tygar/for.i206/1cat3dogs1turtle.txt,133

No contents were found – due to all the URL's being broken.

8. When Ctrl-C is pressed. Program exits with the statement "Program interrupted"
9. Words which does not appear in the text, will give the following output

Search term? hello

The word hello does not appear in any books in the library.

- ## 10. Multiple tabs as query

Search term?

The word does not appear in any books in the library.

- ## 11. Space as query

Search term?

The word does not appear in any books in the library.

12. query with apostrophe – since all the punctuations have been removed during parsing.

Search term? Alice's

The word Alice's does not appear in any books in the library.

13. query with multiple words – words are tokenized during parsing and comparison is for each individual word and not combinations.

Search term? my world

The word my world does not appear in any books in the library.

Search term? this is

The word this is does not appear in any books in the library.

14. query with hyphen

Search term? this-

The word this- does not appear in any books in the library.

Search term? this

The word this appears 493 times in *Pride and Prejudice* (link: [link](#)).

<http://people.ischool.berkeley.edu/~tygar/for.i206/pg1342.txt>)

The word this appears 181 times in Alice's Adventures in Wonderland (link:

<http://people.ischool.berkeley.edu/~tygar/for.i206/pg11.txt>)

The word this appears 76 times in The Yellow Wallpaper (link: <http://people.ischool.berkeley.edu/~tygar/lori2000/pg11.txt>)

The word this appears 15 times in the Yellow Wallpaper (link: <http://people.ischool.berkeley.edu/~tygar/for.i206/pg1952.txt>)

15. Search term? <catalog>

'Pride and Prejudice' : [0, 'http://people.ischool.berkeley.edu/~tygar/for.i206/pg1342.txt']

'Alice's Adventures in Wonderland' : [1, 'http://people.ischool.berkeley.edu/~tygar/for.i206/pg11.txt']

'The Yellow Wallpaper' : [2, 'http://people.ischool.berkeley.edu/~tygar/for.i206/pg1952.txt']

16. Search term? <titles>

Pride and Prejudice

Alice's Adventures in Wonderland

The Yellow Wallpaper

17.

Search term? <terminate>

Program terminated.

For extra credit

The testing was done to ensure that the top 10 ebooks were stripped from the webpage and written into a text file. Since this did not involve any other user input, no further testing was done other than to see if the python program runs properly and gets and writes the required data.