

Project SG2 for Keith Miller @ CS4500

SG2 (Group 5) Developers:

Caleb Jacobs ~ Cynthia Brown ~ DeJa Thompson ~
Dheaven Hill ~ Gabe Jackson

Major Revisions:

10/14/2025 - Joel generated SG0.5

10/14/2025 - SG1 is ready for customer review.

11/09/2025- Initial SG2 Revision

11/13/2025 SG2 revision is ready for customer review.

Table of Contents

1. Overview
2. Test Cases + Files Used
3. SG2 Test Results

Overview

SG2 reads up to ten .txt files, builds individual wordlists, counts total and distinct words, allows users to search across files, and finally generates two reports:

Concordance.txt: every word with its exact (file.line.word) location.

ExtraLists.txt: top frequencies, shared words, and unique words per file.

Test Cases:

Test 1: General Run Through of Program
Test 2: Multiple Files and Duplicate Filenames
Test 3: Hyphenated Line Join Rule
Test 4: Multiple Searches Across Files
Test 5: Concordance file formatted
Test 6: ExtraLists file formatted

Files Used:

- | | |
|--|---|
| <ul style="list-style-type: none">• sample1.txt• sample2.txt• HyphenatedList.txt• test1.txt• test2.txt | <ul style="list-style-type: none">• test3.txt• text4.txt• test5.txt• test6.txt• test7.txt |
|--|---|

SG2 TESTS:

Test 1: General Run Through of Program

Why:

Verify that the program correctly reads a file, counts words, and performs a single search.

Files Used:

sample1.txt

Input:

Enter .TXT filename: *sample1.txt*
Do you want to add another document(yes/no)? *no*
Enter a legal word: *fox*
Would you like to search another word? (Yes/No): *no*

Output:

Filename	TotalWords	Distinct
sample1.txt	27	22

The word "fox" was found in sample1.txt: 2 time(s).

When: 11/09/2025

Who: Gabriel Jackson

What Happened:

Output matched expected totals and occurrences. Concordance and ExtraLists generated successfully.

PASS

Test 2: Multiple Files and Duplicate Filenames

Why:

Ensure SG2 accepts multiple unique files, rejects duplicates, stops at 10 files, and generates a proper summary table.

Files Used:

sample1.txt, sample2.txt, HyphenatedList.txt, test1.txt, test2.txt, test3.txt, test4.txt, test5.txt, test6.txt, test7.txt

Input:

Enter .TXT filename: *sample1.txt*

Add another file? (Yes/No): *yes*

Enter .TXT filename: *sample2.txt*

Add another file? (Yes/No): *yes*

Enter .TXT filename: *sample1.txt*

Enter a new .TXT filename: *HyphenTest.txt*

Enter .TXT filename: *test1.txt*

Add another file? (Yes/No): *yes*

Enter .TXT filename: *test2.txt*

Add another file? (Yes/No): *yes*

Enter .TXT filename: *test3.txt*

Add another file? (Yes/No): *yes*

Enter .TXT filename: *test4.txt*

Add another file? (Yes/No): *yes*

Enter .TXT filename: *test5.txt*

Add another file? (Yes/No): *yes*

Enter .TXT filename: *test6.txt*

Add another file? (Yes/No): *yes*

Enter .TXT filename: *test7.txt*

Add another file? (Yes/No): *yes*

Enter .TXT filename: *test8.txt*

Output:

Filename	TotalWords	Distinct
sample1.txt	27	22
sample2.txt	25	22
HyphenTest.txt	32	29
test2.txt	1329	580
test3.txt	214	130
test4.txt	1363	696
test5.txt	1095	476
test6.txt	259	173
test7.txt	1095	476
test8.txt	1363	696

When: 11/13/2025

Who: Cynthia Brown

What Happened:

Duplicate file correctly blocked.

After 10 files, moved on to word search

Summary table appropriately formatted.

PASS

Test 3: Hyphenated Line Join Rule

Why:

Verify that hyphenated words spanning two lines merge correctly, and stray hyphens are ignored.

File Used: HyphenTest.txt

Input:

Enter .TXT filename: *HyphenTest.txt*

Enter a legal word: *eco-friendly*

Would you like to search another word? (Yes/No): no

Output:

Filename	TotalWords	Distinct
HyphenTest.txt	32	29

The word "eco-friendly" was found in HyphenTest.txt: 1 time(s).

words found: 1

eco-friendly: {'HyphenTest.txt': 1}

Building Concordance

a 1.3.4.

added 1.3.3.

and 1.6.2.

behind 1.2.5.

biodiverse 1.6.3.

broken 1.8.2.

building 1.2.8.

by 1.8.4.

each 1.3.1.

eco-balanced 1.5.6; 1.6.1.

eco-friendly 1.1.2; 1.2.1.

Building Extra Lists.

Top 10 Frequent Words:

Word	Count	Files
-----	-----	-----
the	4	1
eco-balanced	2	1
eco-friendly	2	1
hand-picked	2	1
a	1	1
added	1	1
and	1	1
behind	1	1
biodiverse	1	1
broken	1	1

ensures 1.5.2.

garden 1.2.2.

hand-picked 1.3.5; 1.4.1.

native 1.4.2.

not 1.7.1.

or 1.8.1.

planted 1.2.4.

plot 1.4.6.

project 1.5.4.

random 1.8.5.

science 1.2.7.

species 1.4.3.

stays 1.5.5.

strange 1.7.2.

student 1.3.2.

symbols 1.8.6.

the 1.1.1; 1.2.6; 1.4.5; 1.5.3.

this 1.5.1.

to 1.4.4.

up 1.8.3.

was 1.2.3.

Concordance.txt was created
and written to.

Words that appear in All Files:

a	ensures	species
added	garden	stays
and	hand-picked	strange
behind	native	student
biodiverse	not	symbols
broken	or	the
building	planted	this
by	plot	to
each	project	up
eco-balanced	random	was
eco-friendly	science	

Words that appear in only one file:

Word	File#		
-----	-----		
a	1	eco-friendly	1
added	1	ensures	1
and	1	garden	1
behind	1	hand-picked	1
biodiverse	1	native	1
broken	1	not	1
building	1	or	1
by	1	planted	1
each	1	plot	1
eco-balanced	1	project	1
		random	1
		science	1
		species	1
		stays	1
		strange	1
		student	1
		symbols	1
		the	1
		this	1
		to	1
		up	1
		was	1

When: 11/13/2025

Who: Cynthia Brown

What Happened: Joined words counted once; standalone hyphens ignored, documented in concordance as one word.

PASS

Test 4: Multiple Searches Across Files

Why:

Confirm that multiple searches across several files produce accurate results and session history.

Files Used: sample1.txt, sample2.txt

Input:

Enter .TXT filename: sample1.txt
Add another file? (Yes/No): yes
Enter .TXT filename: sample2.txt
Add another file? (Yes/No): no
Enter a legal word: fox
Would you like to search another word? (Yes/No): y
Enter a legal word: dog
Would you like to search another word? (Yes/No): y
Enter a legal word: testing
Would you like to search another word? (Yes/No): n

Output:

Filename	TotalWords	Distinct
sample1.txt	26	24
sample2.txt	25	22

The word "fox" was found in sample1.txt: 2 time(s).

The word "fox" was found in sample2.txt: 2 time(s).

The word "dog" was found in sample1.txt: 2 time(s).

The word "dog" was found in sample2.txt: 2 time(s).

The word "testing" was found in sample1.txt: 1 time(s).

The word "testing" was found in sample2.txt: 1 time(s).

fox: {'sample1.txt': 2, 'sample2.txt': 2}
dog: {'sample1.txt': 2, 'sample2.txt': 2}
testing: {'sample1.txt': 1, 'sample2.txt': 1}

When: 11/13/2025

Who: Cynthia Brown

What Happened: All search loops worked correctly; summary output matched expectations; program transitioned to SG2 phase successfully.

PASS

Test 5: Concordance file formatted

Why:

Confirm that the concordance is written to a txt file and is formatted correctly.

Files Used: sample1.txt, HyphenTest.txt

Input:

Enter .TXT filename: sample1.txt
Add another file? (Yes/No): yes
Enter .TXT filename: HyphenTest.txt
Add another file? (Yes/No): no

Output:

Concordance.txt Summary:

a 1.3.3; 2.3.4.	fox 1.1.4.	random 2.8.5.
added 2.3.3.	foxs 1.2.9.	science 2.2.7.
amused 1.2.6.	garden 2.2.2.	simple 1.3.4.
and 2.6.2.	hand-picked 2.3.5; 2.4.1.	species 2.4.3.
behind 2.2.5.	however 1.2.3.	stays 2.5.5.
biodiverse 2.6.3.	is 1.3.2.	strange 2.7.2.
broken 2.8.2.	jumps 1.1.5.	student 2.3.2.
brown 1.1.3.	lazy 1.1.8.	symbols 2.8.6.
building 2.2.8.	native 2.4.2.	testing 1.3.8.
by 1.2.7; 2.8.4.	not 1.2.5; 2.7.1.	text 1.3.6.
dog 1.1.9; 1.2.2.	or 2.8.1.	the 1.1.1; 1.1.7; 1.2.1; 1.2.8;
each 2.3.1.	over 1.1.6.	2.1.1; 2.2.6; 2.4.5; 2.5.3.
eco-balanced 2.5.6; 2.6.1.	planted 2.2.4.	this 1.3.1; 2.5.1.
eco-friendly 2.1.2; 2.2.1.	plot 2.4.6.	to 2.4.4.
ensures 2.5.2.	project 2.5.4.	tricks 1.2.10.
example 1.3.5.	purposes 1.3.9.	up 2.8.3.
for 1.3.7.	quick 1.1.2.	was 1.2.4; 2.2.3.

When: 11/13/2025

Who: Cynthia Brown

What Happened: All data was formatted and written to concordance.txt

PASS

Test 5: ExtraLists file formatted

Why:

Confirm that the ExtraLists are written to a txt file and are formatted correctly.

Files Used: sample1.txt, HyphenTest.txt

Input:

Enter .TXT filename: sample1.txt

Add another file? (Yes/No): yes

Enter .TXT filename: HyphenTest.txt

Add another file? (Yes/No): no

Output:

ExtraLists.txt Summary:

Top 10 Frequent Words:

Word	Count	Files
the	8	2
a	2	2
by	2	2
dog	2	1
eco-balanced	2	1
eco-friendly	2	1
hand-picked	2	1
not	2	2
this	2	2
was	2	2

Words that appear in All Files are:

a
by
not
the
this
was

Words that appear in only one file are:

Word	File#					

added	2		fox	1	quick	1
amused	1		foxs	1	random	2
and	2		garden	2	science	2
behind	2		hand-picked	2	simple	1
biodiverse	2		however	1	species	2
broken	2		is	1	stays	2
brown	1		jumps	1	strange	2
building	2		lazy	1	student	2
dog	1		native	2	symbols	2
each	2		or	2	testing	1
eco-balanced	2		over	1	text	1
eco-friendly	2		planted	2	to	2
ensures	2		plot	2	tricks	1
example	1		project	2	up	2
for	1		purposes	1		

When: 11/13/2025

Who: Cynthia Brown

What Happened: All data was formatted and written to concordance.txt

PASS
