

Cerca per Query amb Pesos – Jocs de Prova

Per tots els casos de prova d'aquest document utilitzem un seguit de fitxers de novel·les obtinguts del Projecte Gutenberg, que es llisten a continuació a mode de referència. En tots els casos suposem que ja són carregats al sistema.

Title	Author	Creation Date	Modification Date	File Name
A Manual of Moral Philosophy	Andrew Preston Peabody	2022-12-22T21:02:28.976991	2022-12-22T21:02:29.396401	27531-0.txt
On the Origin of Species 1st Edition	Charles Darwin	2022-12-22T21:02:29.416471	2022-12-22T21:02:30.588579	DarwinOriginofSpecies.txt
A Christmas Carol	Charles Dickens	2022-12-22T21:02:30.613423	2022-12-22T21:02:30.867221	DickensAChristmasCarol.txt
Great Expectations	Charles Dickens	2022-12-22T21:02:30.897580	2022-12-22T21:02:32.138156	DickensGreatExpectations.txt
The Pickwick Papers	Charles Dickens	2022-12-22T21:02:32.172176	2022-12-22T21:02:34.361306	DickensThePickwickPapers.txt
The Jungle Book	Rudyard Kipling	2022-12-22T21:02:34.394443	2022-12-22T21:02:34.765512	KiplingJungleBook.txt
The Call of the Wild	Jack London	2022-12-22T21:02:34.802677	2022-12-22T21:02:35.044558	LondonCallofTheWild.txt
The Commission in Lunacy	Honore de Balzac	2022-12-22T21:02:35.495886	2022-12-22T21:02:35.712420	pg1410.txt
New York	James Fenimore Cooper	2022-12-22T21:02:38.663236	2022-12-22T21:02:38.767873	pg2482.txt
Moonbeams From the Larger Lunacy	Stephen Leacock	2022-12-22T21:02:44.734394	2022-12-22T21:02:45.102197	pg4064.txt
Giordano Bruno	Walter Horatio Pater	2022-12-22T21:02:45.132486	2022-12-22T21:02:45.213523	pg4228.txt
The Sea Fairies	L. Frank Baum	2022-12-22T21:02:46.726128	2022-12-22T21:02:47.020553	pg4358.txt
The Memoirs of General W. T. Sherman	William T. Sherman	2022-12-22T21:02:47.049043	2022-12-22T21:02:49.568337	pg4361.txt
The Vampyre; A Tale	John William Polidori	2022-12-22T21:02:49.596291	2022-12-22T21:02:49.702780	pg6087.txt
Barford Abbey	Susannah Minific Gunning	2022-12-22T21:02:35.073007	2022-12-22T21:02:35.478418	pg13314.txt
A History of Trade Unionism in the U.S.	Selig Perlman	2022-12-22T21:02:35.747556	2022-12-22T21:02:36.392826	pg14458.txt
History Of Egypt, ChaldaËa, Syria, Babylonia, Assyria	G. Maspero	2022-12-22T21:02:36.418815	2022-12-22T21:02:37.430122	pg17326.txt
Traditions of the North American Indians	James Athearn Jones	2022-12-22T21:02:37.460644	2022-12-22T21:02:38.060971	pg20826.txt
The Song of the Exile--A Canadian Epic	Wilfred S. Skeats	2022-12-22T21:02:38.079250	2022-12-22T21:02:38.237201	pg20939.txt
The Emancipation Proclamation	Abraham Lincoln	2022-12-22T21:02:38.282029	2022-12-22T21:02:38.313296	pg22082.txt
Dorothy and the Wizard in Oz	L. Frank Baum	2022-12-22T21:02:38.348663	2022-12-22T21:02:38.643208	pg22566.txt
When Ghost Meets Ghost	William Frend De Morgan	2022-12-22T21:02:38.785765	2022-12-22T21:02:41.305709	pg30896.txt
The English Novel in the Time of Shakespeare	J. J. Jusserand	2022-12-22T21:02:41.323281	2022-12-22T21:02:42.074115	pg31151.txt
Logic, Inductive and Deductive	William Minto	2022-12-22T21:02:42.098500	2022-12-22T21:02:42.782963	pg31796.txt
Great Musical Composers: German, French, Italian, English	George T. Ferris	2022-12-22T21:02:42.815283	2022-12-22T21:02:43.763038	pg34381.txt
Strange Stories from the Lodge of Leisler	Unknown	2022-12-22T21:02:43.789120	2022-12-22T21:02:44.018135	pg37766.txt
The Code of the Mountains	Charles Neville Buck	2022-12-22T21:02:44.055062	2022-12-22T21:02:44.706949	pg38498.txt
The Bible Of Bibles - Or Twenty-Seven Books	Kersey Graves	2022-12-22T21:02:45.265025	2022-12-22T21:02:46.700855	pg43550.txt
The Works of Edgar Allan Poe - Vol. 1	Edgar Allan Poe	2022-12-22T21:02:49.729225	2022-12-22T21:02:50.371486	PoeWorksVol1.txt
The Works of Edgar Allan Poe - Vol. 2	Edgar Allan Poe	2022-12-22T21:02:50.400588	2022-12-22T21:02:51.068365	PoeWorksVol2.txt
A Princess of Mars	Edgar Rice Burroughs	2022-12-22T21:02:51.102132	2022-12-22T21:02:51.616355	RiceBurroughsAPrincessofMars.txt
The Time Machine	H. G. Wells	2022-12-22T21:02:51.655745	2022-12-22T21:02:51.951742	WellsTimeMachine.txt
The War of the Worlds	H. G. Wells	2022-12-22T21:02:51.994789	2022-12-22T21:02:52.450337	WellsWarofTheWorlds.txt

Cas Estàndard

Objectius

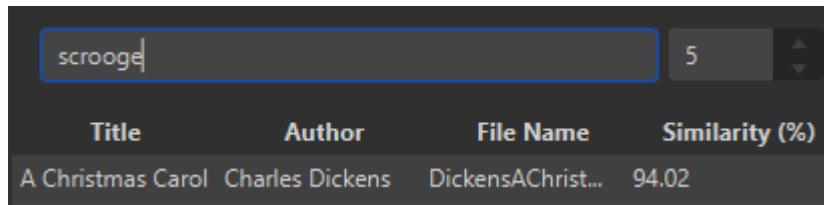
Volem comprovar que els resultats obtinguts en cercar documents amb una query amb pesos són raonables – és clar, es tracta d'una cerca aproximada, pel que ens conformem amb resultats que “tinguin sentit”.

És esperable que, en cercar paraules prominents en la trama o el tema d'una novel·la, el resultat més similar sigui la novel·la de què les hem extret.

Passos a seguir

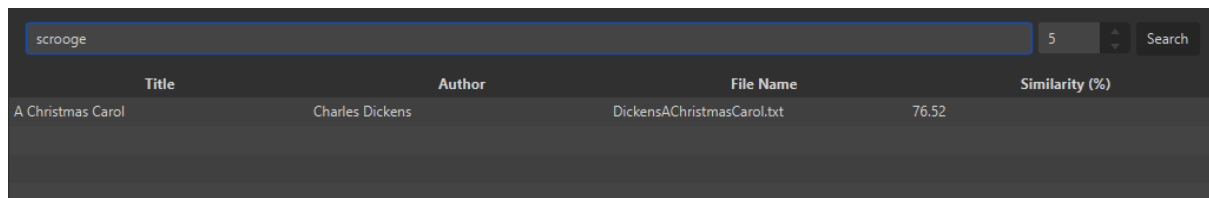
Utilitzarem com a primer exemple “A Christmas Carol”, de Charles Dickens. El personatge principal és el senyor Scrooge: si cerquem aquesta paraula, esperem que (1) sigui l'únic resultat i (2) tingui un nivell de semblança elevat.

Per començar, premem la icona de la balança (per “Weighted search”) i se'ns obre el següent panell.



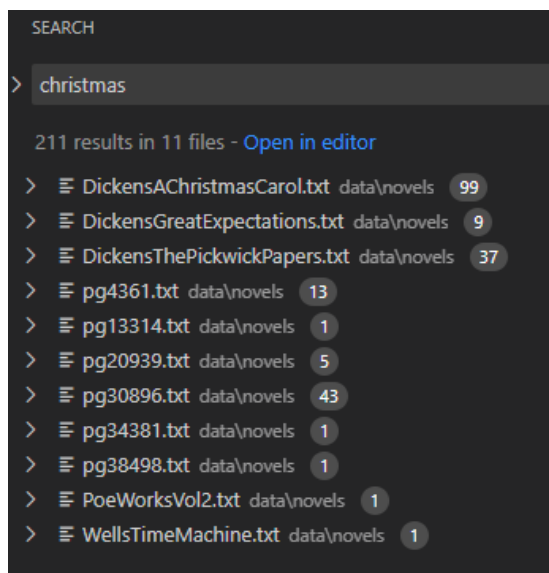
Title	Author	File Name	Similarity (%)
A Christmas Carol	Charles Dickens	DickensAChrist...	94.02

En efecte, quan introduïm la paraula “scrooge”, l'únic resultat és “A Christmas Carol”, amb una semblança molt elevada, del 94%.

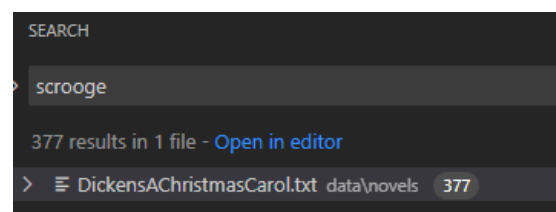


Title	Author	File Name	Similarity (%)
A Christmas Carol	Charles Dickens	DickensAChristmasCarol.txt	76.52

A continuació, modifiquem la query afegint la paraula “christmas”. Per comprovar que els resultats tenen sentit, cerquem el nombre d'ocurrències de “scrooge” i de “christmas”, utilitzant la funcionalitat de cerca de Visual Studio Code.



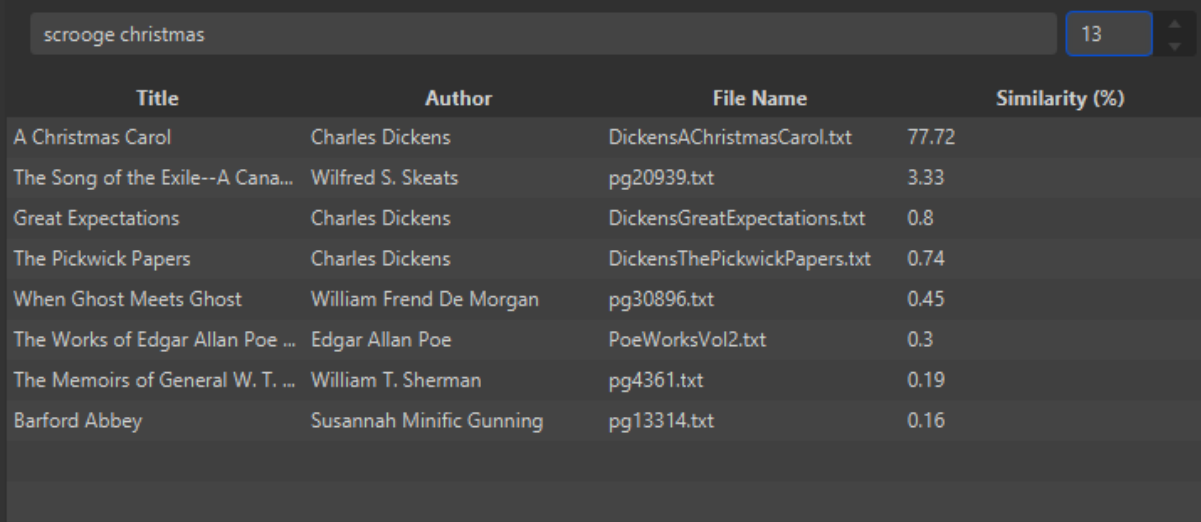
File Name	Occurrences
DickensAChristmasCarol.txt	99
DickensGreatExpectations.txt	9
DickensThePickwickPapers.txt	37
pg4361.txt	13
pg13314.txt	1
pg20939.txt	5
pg30896.txt	43
pg34381.txt	1
pg38498.txt	1
PoeWorksVol2.txt	1
WellsTimeMachine.txt	1



File Name	Occurrences
DickensAChristmasCarol.txt	377

Veiem que “scrooge” apareix moltes vegades, així com “christmas”, a més, apareix en d'altres novel·les. En conseqüència, esperem (1) que les altres novel·les apareguin al

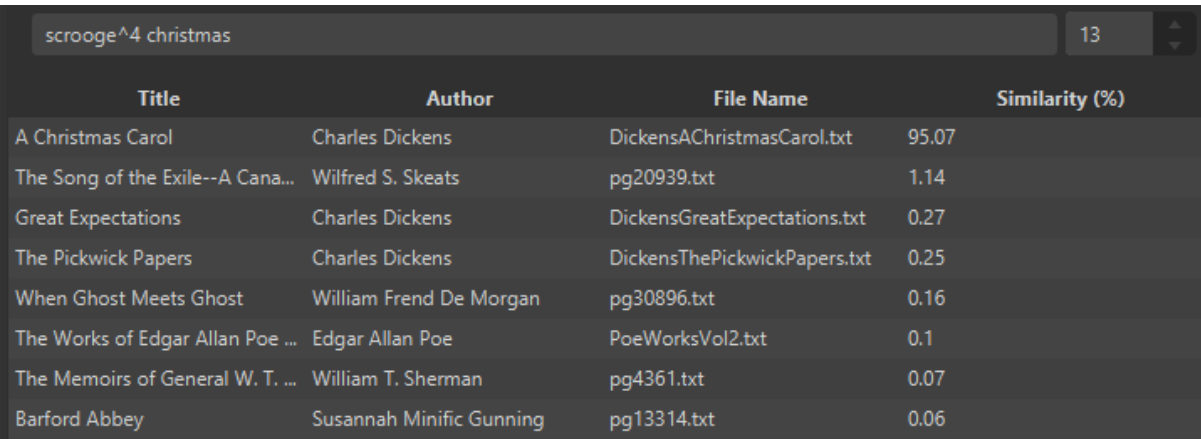
resultat i (2) que la semblança de “A Christmas Carol” sigui menor.



The screenshot shows a search bar with the text 'scrooge christmas' and a dropdown menu showing '13' results. Below the search bar is a table with four columns: Title, Author, File Name, and Similarity (%).

Title	Author	File Name	Similarity (%)
A Christmas Carol	Charles Dickens	DickensAChristmasCarol.txt	77.72
The Song of the Exile--A Cana...	Wilfred S. Skeats	pg20939.txt	3.33
Great Expectations	Charles Dickens	DickensGreatExpectations.txt	0.8
The Pickwick Papers	Charles Dickens	DickensThePickwickPapers.txt	0.74
When Ghost Meets Ghost	William Frend De Morgan	pg30896.txt	0.45
The Works of Edgar Allan Poe ...	Edgar Allan Poe	PoeWorksVol2.txt	0.3
The Memoirs of General W. T. ...	William T. Sherman	pg4361.txt	0.19
Barford Abbey	Susannah Minific Gunning	pg13314.txt	0.16

Ara, si donem més importància al mot “scrooge” utilitzant els pesos, veurem que la similitud amb “A Christmas Carol augmentarà” (i, és clar, disminuirà amb les altres).



The screenshot shows a search bar with the text 'scrooge^4 christmas' and a dropdown menu showing '13' results. Below the search bar is a table with four columns: Title, Author, File Name, and Similarity (%).

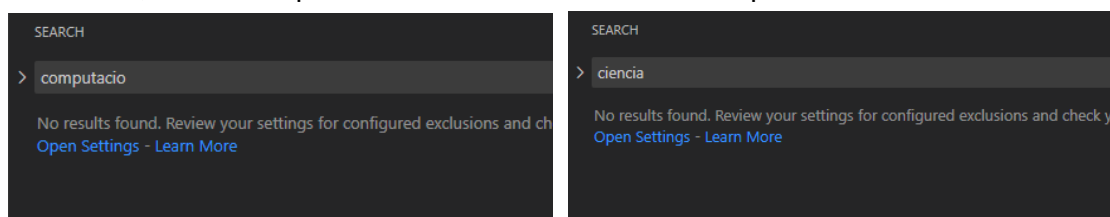
Title	Author	File Name	Similarity (%)
A Christmas Carol	Charles Dickens	DickensAChristmasCarol.txt	95.07
The Song of the Exile--A Cana...	Wilfred S. Skeats	pg20939.txt	1.14
Great Expectations	Charles Dickens	DickensGreatExpectations.txt	0.27
The Pickwick Papers	Charles Dickens	DickensThePickwickPapers.txt	0.25
When Ghost Meets Ghost	William Frend De Morgan	pg30896.txt	0.16
The Works of Edgar Allan Poe ...	Edgar Allan Poe	PoeWorksVol2.txt	0.1
The Memoirs of General W. T. ...	William T. Sherman	pg4361.txt	0.07
Barford Abbey	Susannah Minific Gunning	pg13314.txt	0.06

Cas Extrem – Cerca Completament Diferent

Comprovem que cercar els documents semblants a una query que conté paraules que no apareguin en cap document no retorna cap resultat

Passos a seguir

Cerquem (amb VSCode), algunes paraules que no apareguin en cap novel·la. Per facilitar-ho, escollirem paraules en català: “ciencia” i “computacio”



Veiem que, en efecte, cercar aquests dos mots no retorna cap resultat.

5

No matches

Title	Author	File Name	Similarity ...
-------	--------	-----------	----------------

Queries Incorrectes

Objectius

Volem comprovar que, en introduir-se una query sintàcticament incorrecta, no s'obté cap resultat i es descriu l'error correctament.

Passos a seguir

Escrivim uns quants casos d'expressions incorrectes i comprovem que, en efecte, apareix l'error corresponent.

5

ERROR: Unexpected incomplete '^' found in scrooge^

Title	Author	File Name	Similarity ...
-------	--------	-----------	----------------

5

ERROR: Unexpected incomplete '^' found in ^2

5

ERROR: Only one '^' expected in scrooge^^2

5

ERROR: Unexpected non-alphanumeric token '.5'

5

ERROR: Invalid weight 'dos'

5

ERROR: Empty query