csv (Comma Separated Value) Files

What is a csv file?

A csv file is a plain text formatted file in which values are separated by commas (Comma Separated Values or csv).

How can I open a csv file?

As a plain text file, you can open a csv file with any text editors (e.g., Notepad, Notepad++, TextEdit, VScode). But… the opened file will look a little messy, quite unreadable with the information in the various (comma separated) fields all bunched together.

The CoNLL table of *The Three Little Pigs* file will look like this when openeded in Notepad++

```
ID,Form,Lemma,POStag,NER,Head,DepRel,Clause Tag,Record ID,Sentence ID,Document ID,Document
1,THERE,there,RB,O,2,advmod,S,1,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
2,was,be,VBD,O,0,ROOT,VP,2,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
3,an,a,DT,O,4,det,NP,3,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
4,old,old,JJ,O,5,nsubj,O,4,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
5,sow,sow,VBP,O,2,ccomp,VP,5,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
6,with,with,IN,O,9,case,0,6,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
7,three,three,CD,NUMBER,9,nummod,NP,7,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
8,little,little,JJ,O,9,amod,0,8,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
9,pigs,pig,NNS,O,5,obl:with,0,9,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
10,",",",",",",O,2,punct,0,10,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
11,and,and,CC,O,22,cc,0,11,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
12,as,as,IN,O,14,mark,SBAR,12,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
13,she,she,PRP,O,14,nsubj,NP,13,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
14,had,have,VBD,O,22,advcl:as,VP,14,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
15,not,not,RB,O,16,advmod,NP,15,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
16,enough,enough,JJ,O,14,obj,0,16,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
17,to,to,TO,O,18,mark,VP,17,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
18,keep,keep,VB,O,16,dep,VP,18,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
19,them,they,PRP,O,18,obj,NP,19,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
20,",",",",",",O,22,punct,0,20,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
21,she,she,PRP,O,22,nsubj,0,21,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
22,sent,send,VBD,O,2,conj:and,VP,22,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
23,them,they,PRP,O,22,obj,0,23,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
24,out,out,RP,O,22,compound:prt,0,24,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
25,to,to,TO,O,26,mark,VP,25,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
26,seek,seek,VB,O,22,advcl:to,VP,26,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
27,their,they,PRP$,O,28,nmod:poss,NP,27,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
28,fortune,fortune,NN,O,26,obj,0,28,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
29,.,.,.,.,O,2,punct,0,29,1,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
1,The,the,DT,O,2,det,NP,30,2,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
2,first,first,JJ,ORDINAL,16,nsubj,0,31,2,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
3,that,that,WDT,O,2,ref,SBAR,32,2,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
4,went,go,VBD,O,2,acl:relcl,VP,33,2,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
5,off,off,RP,O,4,compound:prt,0,34,2,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
6,met,meet,VBD,O,0,ROOT,VP,35,2,1,"=hyperlink(""C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt"")"
```

The same file opened in Excel is far more readable.

ID	Form	Lemma	POStag	NER	Head	DepRel	Clause Tag	Record ID	Sentence ID	Document ID	Document
1	THERE	there	RB	O	2	advmod	S	1	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
2	was	be	VBD	O	0	ROOT	VP	2	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
3	an	a	DT	O	4	det	NP	3	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
4	old	old	JJ	O	5	nsubj	0	4	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
5	sow	sow	VBP	O	2	ccomp	VP	5	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
6	with	with	IN	O	9	case	0	6	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
7	three	three	CD	NUMBER	9	nummod	NP	7	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
8	little	little	JJ	O	9	amod	0	8	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
9	pigs	pig	NNS	O	5	obl:with	0	9	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
10	,	,	,	O	2	punct	0	10	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
11	and	and	CC	O	22	cc	0	11	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
12	as	as	IN	O	14	mark	SBAR	12	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
13	she	she	PRP	O	14	nsubj	NP	13	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
14	had	have	VBD	O	22	advcl:as	VP	14	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
15	not	not	RB	O	16	advmod	NP	15	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
16	enough	enough	JJ	O	14	obj	0	16	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
17	to	to	TO	O	18	mark	VP	17	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
18	keep	keep	VB	O	16	dep	VP	18	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
19	them	they	PRP	O	18	obj	NP	19	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt
20	,	,	,	O	22	punct	0	20	1	1	C:/Program Files (x86)/NLP/lib/sampleData/The Three Little Pigs.txt

Opening a csv file in Excel (recommended) or Numbers

Windows

Follow the steps below to ensure that Microsoft Excel is set as your default program for CSV files:

1. Start >Default Programs.
2. Click Associate a file type or protocol with a program.
3. Locate the CSV option and click it once to highlight.
4. Click Change program.
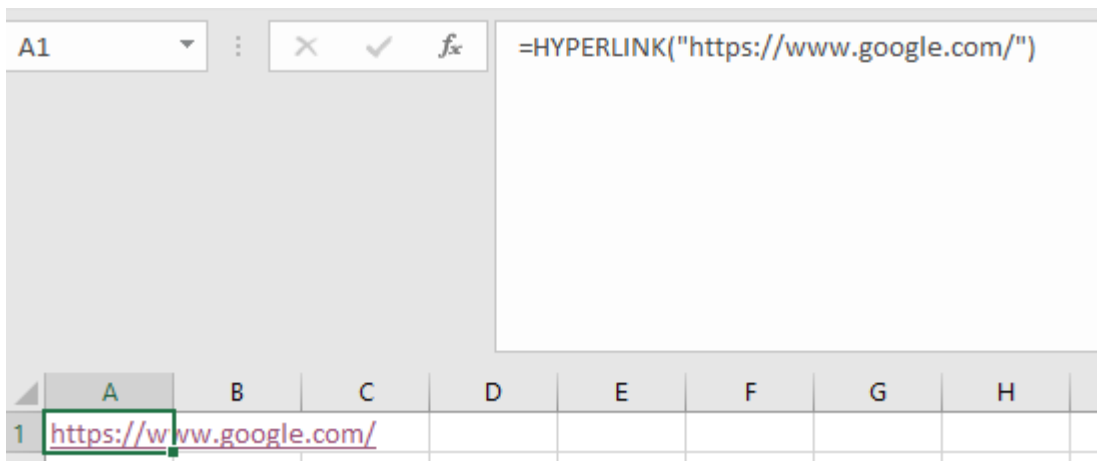5. Click the Microsoft Excel option under Recommended Programs.
6. Click OK.

Mac

1. Right-click on an Excel File (. xlsx, . xls, . csv, etc)
2. Click on Get Info in the pop-up.
3. Click Open With if the section isn't already expanded.
4. Click the dropdown and choose Microsoft Excel.
5. Click Change All...
6. Click Continue when the Finder asks you if you're sure.

What is the difference between a csv file and an Excel file?

The difference between CSV and XLS file formats is that CSV format is a plain text format in which values are separated by commas (Comma Separated Values), while XLS file format is an Excel Sheets binary file format which holds information about all the worksheets in a file, including both content and formatting.

Hyperlinks: What are they?

A hyperlink is a word, phrase, or image that you can click on to jump to a new document or a new section within the current document, or to a webpage for that matter. Hyperlinks are found in nearly all Web pages, allowing users to click their way from page to page.

In a csv file opened in Excel, when you select a cell with a hyperlinked item, first, the hyperlink will be displayed on top with the value =HYPERLINK… If you click on cell A1, the hyperlink will open the Google page.

Weird characters in my csv file opened in Mac Numbers: Why do I get this?

The default spreadsheet used in Mac is Numbers. But… Mac blocks hyperlinks. If the hyperlink looks like a red warning triangle… You have opened the csv file with Mac Numbers and hyperlinks have been blocked.

Document ID	Document
1	⚠️
2	⚠️
3	⚠️
4	⚠️
5	⚠️
6	⚠️
7	⚠️
8	⚠️
9	⚠️
10	⚠️

You should really use Excel to open the NLP Suite csv files

Many NLP Suite csv output files contain hyperlinks to the documents processed and their location on your computer. In Excel, by clicking on them, you will be able to read the file immediately for accuracy of results, saving you much grief. You would be prevented from doing this in Mac Numbers.

Weird characters in my csv file opened in Excel: Why do I get this?

When I open a csv file created by an NLP Suite Python script (e.g., n-grams) I get some weird characters, such as â€œ on my Windows laptop (not on a Mac). Such characters are not in the text file processed by the script in input.

	A	B	C	D	E	F	G	H
1	4-grams	Frequency	Document	FileName				
2	â€œ What we need	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
3	What we need to	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
4	we need to know	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
5	need to know is	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
6	to know is how	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
7	know is how to	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
8	is how to live	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
9	how to live a	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
10	to live a life	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
11	live a life to	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
12	a life to make	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
13	life to make it	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
14	to make it the	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
15	make it the best	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
16	it the best possible.	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
17	the best possible. â€	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				
18	best possible. â€ â€"	1		1 /Users/Myself/Desktop/NLP/test_input/1.txt				

Oddly, when the same csv file is used to create an Excel chart, the weird character disappears

	A
1	4-grams
2	" What we need
3	What we need to
4	we need to know
5	need to know is
6	to know is how
7	know is how to
8	is how to live
9	how to live a
10	to live a life
11	live a life to
12	a life to make
13	life to make it
14	to make it the
15	make it the best
16	it the best possible.
17	the best possible. "
18	best possible. " –

And the weird character is not in the input text file either.

```
"What we need to know is how to live a life to make it the best possible."

-   Socrate
⌐ And this applies to everyone, in every culture and every country. I am very fortunate to live in a city where people are open-minded and aware of the rights and
freedoms of everyone. In Montreal, being gay is pretty well accepted, even very well accepted. There is very little discrimination and the gay community is very
present. I would even say that here, homophobic are judged more harshly than homosexuals!
Yet even here, to "come out" is not always simple. I came out at the age of 20. Today, in retrospect, I wonder why I waited so long.
```

The weird character â€œ seems to be a csv display of the character ".

while quotation marks and apostrophes seem innocuous, they can cause problems when the CSV is opened, as opposed to imported into Excel, partly because quotation marks or other symbols such as apostrophes come in different forms. Some are curly ", some are straight ". It can depend on what application generated the characters to begin with. It can be hard to explain exactly why Excel has an issue with some characters but read on for some solutions. A **solution** is to replace all curly quote marks to straight ones before saving a csv file. (https://help.shotfarm.com/hc/en-us/articles/115004652968-Why-are-there-odd-characters-when-I-open-a-CSV-in-Excel-)