

Excel Tip: Counting the number of items in a multi value cell

Some *iSearch* fields contain multiple values when downloaded, for example, the number of Authors for a Publication, as shown in the example below.

	A	B	C	D	E	F	G
1	PMID	Pub Year	Title	Authors			
2	24173616	2015	Anatomic	Unal, Bengi;Shah, Fulva;Kothari, Janish;Tepper, James M			
3	24417286	2015	Reward pr	Perry, David C;Kramer, Joel H			
4	24512192	2015	NOX2 As a	Diebold, Becky A;Smith, Susan M E;Li, Yang;David Lambeth, J			
5	24623361	2015	Current pe	Wang, Brian;Jin, Kunlin			
6	24684551	2015	Modes of	Paul, Bindu D;Snyder, Solomon H			
7	24700259	2015	Relations	Bryant, Mon S;Rintala, Diana H;Hou, Jyh Gong;Protas, Elizabeth J			
8	24850078	2015	Modeling	Daniel, Guillaume;Moore, Darren J			
9	24855070	2015	Are therap	Kim, Scott Y H;Vries, Raymond De;Parnami, Sonali;Wilson, Renee;Myra Kim, H;Frank, Samuel;Holloway, Robert G;Kie			
10	24863743	2015	Neureguli	Wu, Limin;Walas, Samantha;Leung, Wendy;Sykes, David B;Wu, Jiang;Lo, Eng H;Lok, Josephine			
11	24880787	2015	The physio	Garcia Rill, E;Hyde, J;Kezunovic, N;Urbano, F J;Petersen, E			

There are ways to answer these questions using *iSearch*. The unwind feature available in the .csv download enables the user to create a row for each value in a multi valued cell but this increases the size of the dataset and may not be helpful depending on the analytical question being investigated. There is also a “count of pubs” field in *iSearch: Grants* that that can be included in a download.

The method and formula

The individual values are separated by a ‘;’ so to know how many unique values there are it is possible to count the number of times ‘;’ appears in the cell and add one to the total (for the last value in the list).

There is no simple formula to count the number of times a character appears in a cell, but it is possible using a combination of Excel formulas:

- The LEN formula counts the total number of characters in a cell.
- The SUBSTITUTE formula replaces a specific character in a cell with another character

So counting the number of characters in a cell and comparing that to the count without any “;” (it is replaced by “” i.e. removed) gives the number of times “;” appears in the cell. Note that the last value is not followed by “;” so it is necessary to add one to the result.

Formula 1: =LEN(D2)-LEN(SUBSTITUTE(D2,";",""))+1

This formula works for any field where every record has at least one value (see below, counting the number of PIs on each grant).

E2 : =LEN(D2)-LEN(SUBSTITUTE(D2,",";""))+1

	A	B	C	D	E	F
1	PMID	Pub Year	Title	Authors	Number of authors	
2	24173616	2015	Anatomic	Unal, Bengi;Shah, Fulva;Kothari, Janish;Tepper, James M	4	
3	24417286	2015	Reward pr	Perry, David C;Kramer, Joel H	2	
4	24512192	2015	NOX2 As a	Diebold, Becky A;Smith, Susan M E;Li, Yang;David Lambeth, J	4	
5	24623361	2015	Current pr	Wang, Brian;Jin, Kunlin	2	
6	24684551	2015	Modes of	Paul, Bindu D;Snyder, Solomon H	2	
7	24700259	2015	Relations	Bryant, Mon S;Rintala, Diana H;Hou, Jyh Gong;Protas, Elizabeth J	4	
8	24850078	2015	Modeling	Daniel, Guillaume;Moore, Darren J	2	
9	24855070	2015	Are theraj	Kim, Scott Y H;Vries, Raymond De;Parnami, Sonali;Wilson, Renee;Myra Kim, H;Frank, S	8	
10	24863743	2015	Neureguli	Wu, Limin;Walas, Samantha;Leung, Wendy;Sykes, David B;Wu, Jiang;Lo, Eng H;Lok, Jos	7	
11	24880787	2015	The physis	Garcia Rill, E;Hyde, J;Kezunovic, N;Urbano, F J;Petersen, E	5	
12	24961994	2015	Two-year	Prodoehl, Janey;Rafferty, Miriam R;David, Fabian J;Poon, Cynthia;Vaillancourt, David	10	

The basic version of the formula will not work where there are blank cells, for example counting the number of PMIDs linked to a Grant as the LEN and SUBSTITUTE formula will return 0 and then add 1 to it. To correct the results for blank values, the formula above can be nested within an IF statement:

Formula 2: =IF(I2="",0,LEN(I2)-LEN(SUBSTITUTE(I2,",";""))+1)

So if the cell is empty the count is set to zero, otherwise the formula will count the “;”s and provide a number of values in the cell. This formula is used to calculate the number of PMIDs in the example below.

K2 : =IF(I2="",0,LEN(I2)-LEN(SUBSTITUTE(I2,",";""))+1)

	A	B	C	D	E	F	G	H	I	J	K	L	M
		Grant						Activity					
1	System ID	Number	Appl ID	Title	All PIs	All PIs IDs	Fiscal Year	Code	PMID	Number of PIs	Number of PMIDs		
2	8717189	R01NR014	8717189	Healing Hi	Gary, Rebecca A;Waldrop Valverde, Drenna	2090579;7598287	2015	R01		2	0		
3	8725426	R01NR014	8725426	Reducing	Jones, Rachel	7743015	2015	R01	28528463;27405326	1	2		
4	8770480	R01DA038	8770480	Networks	Wagner, Karla D	8952774	2015	R01		1	0		
5	8789092	R01MH105	8789092	Examining	Weeks, Margaret R	1960789	2015	R01		1	0		
6	8790362	R01MH105	8790362	Cysteinyln	Kaul, Marcus	7019491	2015	R01	26482305;26968848;28241493;27664068	1	4		
7	8838888	R01AI1155	8838888	Timing of	Swanstrom, Ronald I;Williamson, Carolyn	1870560;9177169	2015	R01	26959192;26813340;27064995	2	3		
8	8838891	R01AI1155	8838891	Analysis o	Maskew, Mhairi;Fox, Matthew Alexander	10682480;9560914	2015	R01	28099905;28225717	2	2		

Note that Formula 2 could also be used in place of Formula 1 in the PI example and would give the same result.

Note: Care needs to be taken when using this formula on iSearch downloads and IMPACII data:

- There are limits on the number of characters Excel will accept in a field, so for Grants with thousands of linked publications, the list of PMIDs will be truncated.
- There may be duplicate Authors / PIs linked to a single publication / grant. Counting the number of items in the field would potentially double count and inflate the numbers.
- PMIDs are linked to core grants so counting multiple ApplIDs within a core grant number will potentially lead to double counting.

As with any analysis, quality checking of the underlying data is important.

If you are having problems, contact OPA training: OPA-Training@mail.nih.gov