

## Re: tagging unique items in a list

---

*Source:*

<http://www.tech-archive.net/Archive/Excel/microsoft.public.excel.worksheet.functions/2006-06/msg01859.html>

---

- *From:* K. Gwynn <[KGwynn@xxxxxxxxxxxxxxxxxxxxxxxxxxxx](mailto:KGwynn@xxxxxxxxxxxxxxxxxxxxxxxxxxxx)>
  - *Date:* Fri, 16 Jun 2006 05:58:01 -0700
- 

Actually, I've done that too! I guess I then need to put back into a single column the areas that form the "uniqueness"? I was trying to avoid all those extra steps, but I guess there isn't any way to do that.

"Otto Moehrbach" wrote:

You can see for yourself that the variation in what number character means what makes your explanation too difficult to follow for anyone but you. Do this. Copy the whole column of these entries into Column A of a new sheet starting with A1. Then select all the cells that have entries in this new sheet. Now click on Data – Text to columns. Select delimited. Click the "Other" box and insert the underscore character. Select Text. Click on OK. This puts each piece of each entry into a separate column. Now explain how to determine if a row is unique. Don't get into groups and subgroups. Stick with columns, Column A means this, Column B means that, and so forth. I know that what you explained makes sense to you, but you have to realize that you are in the minority on this. What you want to do is very easy for Excel to do, if I knew what you wanted. Ignore the fact that now all of your entries are distributed throughout a bunch of columns, one piece per column. Excel can put this egg back together again very easily. HTH Otto

"K. Gwynn" <[KGwynn@xxxxxxxxxxxxxxxxxxxxxxxxxxxx](mailto:KGwynn@xxxxxxxxxxxxxxxxxxxxxxxxxxxx)> wrote in message [news:2D520DC9-3B54-41FD-807D-E8DFAC886799@xxxxxxxxxxxxxxxxxxxx](mailto:news:2D520DC9-3B54-41FD-807D-E8DFAC886799@xxxxxxxxxxxxxxxxxxxx)

okay: the entries in my data set can have anywhere between 17 and 27 characters.

The entries are broken into character strings by the underscore character. Two groups of character strings are important: those that identify the entry as belonging to a particular subgroup and those that then determine whether within that subgroup that entry is unique.

Each subgroup where we determine if one entry is different from another is identified by the characters in positions 7 through 12 or 7 through 14.

The subgroup that determines whether an entry is unique within that subgroup

Re: tagging unique items in a list

is  
either the last character or starts 8 to 9 characters from the right.

Here's another example:

- (1) CMMI2\_MA\_SP2\_04\_P\_ND\_PP\_H
- (2) CMMI2\_PMC\_GP2\_01\_G
- (3) CMMI2\_PPQA\_GP2\_01\_P
- (4) CMMI2\_PPQA\_GP2\_01\_P\_PD\_PP\_H

In this example there are three subgroups: line 1; line 2; line 3 & 4 each represent unique subgroups. Because lines 1 & 2 only have one entry in their respective subgroups, they are both unique. The last 8 characters of line 4 (and really line 1) are always superfluous. Therefore, to determine if entries 3 & 4 are duplicates we move past those characters and look at the last character in line 3 and what would be the last character in line 4 if the group of 8 "extra" characters were removed from line 4. When viewed in that way, we see that these entries are duplicates and we only need to keep one entry (doesn't matter which). We now have three unique entries to move to our new column:

- (1) CMMI2\_MA\_SP2\_04\_P\_ND\_PP\_H
- (2) CMMI2\_PMC\_GP2\_01\_G
- (3) CMMI2\_PPQA\_GP2\_01\_P

I want to copy the unique set to a new column, not move. I need to keep the original set of data.

"Otto Moehrbach" wrote:

You say the second set of characters can have different lengths. What about the first set, the set of 17 characters in your example? Also, you say you want the unique entries "moved". Not copied but moved?  
Otto  
"K. Gwynn" <KGwynn@xxxxxxxxxxxxxxxxxxxxxxxxxxxx>  
wrote in message  
[news:56A61AB8-70C4-495A-B577-B1C8C68C39BE@xxxxxxxxxxxxxxxxxxxx](mailto:news:56A61AB8-70C4-495A-B577-B1C8C68C39BE@xxxxxxxxxxxxxxxxxxxx)

Yes, I am interested in only finding the entries that are unique. But its not always the same number of characters. The second set of characters is

Re: tagging unique items in a list

sometimes as many as four characters long.  
The stuff at the end could  
be  
as  
short as 3 characters or as long as 5 extra  
characters.  
Once identified, I need to have the unique  
entries moved to another  
column  
so I can work just with them.

"Otto Moehrbach" wrote:

Correct me if I'm wrong  
here, but it appears that you  
are not  
interested  
in  
finding duplicate entries, but  
only those entries that are  
duplicate  
for  
the  
first specific number of  
characters. In your examples  
that's 17  
characters.

In other words, not counting  
what you refer to as the  
extension. Is  
that  
right?

You would need VBA for  
that but you don't say what  
you want done with  
entries that are found to be  
duplicates (for the first 17  
characters).

Do  
you want duplicates  
highlighted? Moved?  
Copied? All the entries  
that  
are  
duplicates or do you want to  
leave one? HTH Otto

"K. Gwynn"

<KGwynn@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx>

wrote in message

[news:28750A14-DA8E-4FFD-983D-15593877038D@xxxxxxxxxxxxxxxxxxxxx](mailto:news:28750A14-DA8E-4FFD-983D-15593877038D@xxxxxxxxxxxxxxxxxxxxx)

Re: tagging unique items in a list

Before  
posting this  
message I  
read  
through all  
the posts for  
finding  
and  
counting  
unique  
entries and  
tried most  
of the  
responses  
and still  
have  
not  
found a  
solution that  
works for  
my  
situation.

So here  
goes:

I have a  
spreadsheet  
with 700+  
entries. The  
records I  
need to sort  
on  
look  
like this:

CMMI2\_CM\_GP2\_08\_G  
CMMI2\_CM\_GP2\_09\_P  
CMMI2\_CM\_GP2\_09\_P  
CMMI2\_CM\_GP2\_09\_P\_PD\_PP\_H  
CMMI2\_CM\_GP2\_10\_P\_ND\_PP\_H

In the  
example  
above the  
duplicates  
are

CMMI2\_CM\_GP2\_09\_P  
CMMI2\_CM\_GP2\_09\_P  
CMMI2\_CM\_GP2\_09\_P\_PD\_PP\_H

So when

Re: tagging unique items in a list

I've tried  
using  
Advanced  
Filter, the  
two that  
look exactly  
the  
same  
are filtered  
out, but the  
one with the  
"PD\_PP\_H"  
extension is  
not.

[No, that  
extension  
cannot be  
filtered out  
ahead of  
time and  
yes, it  
is  
important to  
the data  
set].

Now what?  
Thanks in  
advance for  
any  
suggestions  
and help.