

Re: How to sort/update large excel db

Source: <http://www.tech-archive.net/Archive/Excel/microsoft.public.excel.misc/2004-10/1117.html>

From: ConfusedNovice (*anonymous_at_discussions.microsoft.com*)

Date: 10/04/04

Date: Mon, 4 Oct 2004 00:53:24 -0700

Thanks a lot. I'll find out how the db repairs worked out sometime soon.

For a unique key value, I am thinking of date & incrementing record number: "100404-1" Only 1 person is likely to be entering data at any given time, but it would be easy to add a terminal designation if needed. There are *many* records that have blank cells in various fields, & I haven't been able to come up with any other ideas for unique values.

About "slipping," maybe I should try to illustrate:

1)Older version of db:

Last First Company Address

Bush GW 1600 Penn

Doe 123 Main St

Co & Co 25 Oak St

Jones John 111 Elm St

Jones Mary 456 Main St

2)Later version of db:

Bush GW 1600 Penn

Doe John 123 Main St

Co & Co 25 Oak St

Jones Pedro 111 Elm St

Jones Mary 456 Main St

Sometimes the shift seems to involve several nearby records & simple fill-in's of blank cells, but other times it's not that straightforward. Some blanks were preserved, some weren't. Some records were ok, some weren't. I've been doing a lot of hand comparisons among whatever db versions & paper records I could get. Just guessing, I'd say serious problems could have been introduced on any or all of 3 monthly updates. Lots of reconstruction, very tedious.

A few yrs ago I was an Access user, but I've never set it up for something like this. Seems like it might be harder to break. How could it be any worse than what I've got now with Excel? Only other db experience = Unidata, never did anything more than simple math spreadsheets in Excel.

Jane

>-----Original Message-----
>I'm not sure I understand the slipping into empty cells stuff.
>
>But I find it better to use a field that must be entered in the list as the
>key--if you make an arbitrary key value (some numbering system that only you
>know is important), then it won't be updated (correctly) by the users--well, I
>wouldn't!
>
>But by using a field that must be completed--even if it is a concatenation of
>cells to make it unique, it makes more sense to most users--and if it's used,
>there's a better chance of typing the values correctly.
>
>(That won't always happen, but hopefully it'll be small potatoes--not a giant
>problem.)
>
>And as long as you have that key, you can sort/filter by anything you want. And
>just resort by that key.
>
>And I don't speak the Access, but if you really want multiple people to update
>records, you may want to consider doing it in Access (or a different real
>database program). From what little I understand, you're forced to use a key
>and it allows simultaneous edits (to different records (I think).
>
>=====
>One last note. Save often. And do File/SaveAs often. You may find that you
>want to go back a couple of updates if you screw something up. And I bet you're
>making changes at a furious rate.
>
>(Disk space is cheap--starting over is expensive!)

>

>

>

>ConfusedNovice wrote:

>>

>> Thanks, Dave. This thread is getting unwieldy, but I

>> **think** I understand what you're advising.

>>

>> Sticking with the issues at hand, yes, the records are all

>> one per row. Last Name is probably the best key, but there

>> are some records with only a Company Name, & a few with

>> only an email name. I have been using Last Name & Zip

>> Codes to try & match things up.

>>

>> I discovered that when sorting some of the intermediate db

>> versions by "new" entries, those newest entries were

>> mostly correct. I am thinking of isolating these records &

>> copy/pasting them on to the oldest db version. The idea is

>> to try for a cleaner sheet to compare to the current one.

>> Maybe it would be better not to merge these? It's crunch

>> time – I will try your formulae on a test file today.

>>

>> And moving ahead: If I can execute the comparisons you

>> suggest, and if I can achieve a cleaner list, then

what's

>> the best way to prevent data from slipping into the empty

>> cells in the future? Should I make a new index column,

>> number each record, and append & number all new records?

>> Can people safely sort & filter the data? What if they

>> update a record in a filtered list & then re-sort? Or

>> Find/Replace & then re-sort? I still don't understand

how

>> some of the data slipped out of one record & into another.

>> Until I do understand, I'm afraid it will continue to

>> happen & all this work will be for nothing.

>>

>> Thanks again! It's an ugly way to learn Excel.

>>

>> Jane

>>

>> Subject: Re: How to sort/update large excel db

>> From: "Dave Peterson" <ec35720@msn.com>

>> Sent: 9/30/2004 1:05:46 PM

>>
>> *Just to clarify, is your data laid out one record per*
>> *row? You don't use*
>> *multiple rows to represent one "logical" record, do you?*
>>
>> *If you do, then this will make it even more difficult.*
>>
>> *But to match/merge two different worksheets, I would*
think
>> *that you'd need to*
>> *have a unique key/index into the data.*
>>
>> *If you don't have this key, how do you match up your*
data
>> *to look for*
>> *differences?*
>>
>> *I wouldn't put too much faith in having that unique key*
in
>> *the last name*
>> *column--but maybe you can use a couple of columns that*
>> *could serve as that key.*
>>
>> *For instance, if you had the lastname, firstname,*
address
>> *in 3 separate columns,*
>> *you could use a helper cell and do something like:*
>> *=A2&char(10)&b2&char(10)&c2*
>> *(The char(10) probably won't appear in your data and*
will
>> *serve as a field*
>> *delimiter--just in case there's records that might look*
>> *identical when*
>> *combined.)*
>>
>> *But even if you did this, you'd have to make sure that*
>> *these fields weren't*
>> *changed. If any were changed, then you couldn't use*
them
>> *for a match between*
>> *worksheets.*
>>
>> *To find out if the records in one worksheet are in the*
>> *other (based on any*
>> *single column you can use--in your data or derived from*
>> *your data):*
>>
>> *=isnumber(match(a2,sheet1!a:a,0))*
>> *and drag down.*
>> *You'll see True if it appears on that other worksheet.*
>>

>> *And use the equivalent to check the other worksheet.*
>>
>> *Once you've found your Falses--you'll want to inspect them*
>> *to see if they should*
>> *be cleansed--cleaned up to make them match.*
>>
>> *You could end up fixing the data on either worksheet--just*
>> *to find the matches.*
>>
>> *Once you get rid of all the Falses, you can do the =vlookup*
>> *() stuff to match*
>> *merge.*
>>
>> *I'm not sure if all this work can be done on a quick*
>> *schedule. This kind of*
>> *stuff usually means fix, check, fix, check....until you*
>> *can't find any more*
>> *differences. Then you let someone else review it and it*
>> *starts again.*
>>
>> *Good luck,*
>>
>> *ConfusedNovice wrote:*
>> >
>> > *Still trying to understand what I've got. There is no*
>> > *column with all unique entries. Every column has blank*
>> > *cells someplace. On 2 of the earlier database*
versions,
>> > *the leftmost column (Column A) is set up as "Index*
>> > *Column," but not all the records have an "Index"*
number
>> *in*
>> > *this column - some of them are blank, & some have*
zeros
>> *in*
>> > *Column A. The most recent master does not have Index*
>> > *column, it starts with a Last Name field.*
>> >
>> > *This brings me back to the question of how to prevent*
>> > *names/values from slipping into the blank cells when*
>> > *sorting & updating records. I have been reading a lot*
of
>> > *Excel info, but this is *not* my area at all, & I*
want to
>> > *fix these problems, not compound them. Should I be*
able
>> *to*
>> > *identify key column(s)? If yes, how? And if no, is*

there

>>> *any safe way to sort when blanks exist in every column?*

>>> *(Or can it be a Key & contain multiple blank cells?)*

>>>

>>> *There are ~17,000 records, & when the wrong values move*

>>> *into a blank cell, that record doesn't indicate that it*

>>> *has been changed. (Probably because it was not opened?)*

>>>

>>> *I'm seriously running out of time. I've got a pretty big*

>>> *list of changes ready to enter directly on db records. I*

>>> *also have isolated a few hundred recent entries by date,*

>>> *verified them, & placed them in a separate worksheet.*

>>> *(Hoping to append to a clean main list & sort into place.)*

>>> *I don't think it's possible to verify every row.*

>>>

>>> *Yes, no matter how you look at this, it's a pain! I'm not*

>>> *really dumb, just a total rookie, & your help is*

>>> *invaluable.*

>>> *To summarize:*

>>> *Whatsup with "Index Column"?*

>>> *How to identify/designate Key columns?*

>>> *Implications of having blank cells throughout sheet?*

>>> *Fastest way to add/edit records if I go back to an older,*

>>> *cleaner version of the database?*

>>>

>>> *Thanks again!*

>>>

>>> *Jane*

>>>

>>> *>-----Original Message-----*

>>> *>I'm saying that if you have two "live" copies of the workbook, it'll be a pain*

>>> *>to merge one of them back into the other.*

>>>>

>>> *>There's nothing built into excel that will force you to*

>>> *make sure your key*

>>> *>column entries are unique.*

>>>>

>>> *>When I do this, I spend some time verifying that the data*

>> > *is at least*
>> > *>consistent--no additional entries with the same key*
were
>> > *added. If you can*
>> > *>believe/verify this, then, yeah, you can do =vlookup*
(*'s*
>> > *to merge the updated*
>> > *>values into the one real workbook.*
>> > >
>> > *>It's just in my experience, avoiding this is usually*
>> > *better. (Just give the*
>> > *>file to someone else and tell them to return it when*
>> > *they're done--and put a*
>> > *>freeze on any other changes.)*
>> > >
>> > >
>
>--
>
>*Dave Peterson*
>*ec35720@msn.com*
>.
>