
Subject: Re: How can I setup right collating sequence in GridCtrl?

Posted by [mr_ped](#) on Wed, 06 May 2009 08:39:24 GMT

[View Forum Message](#) <> [Reply to Message](#)

sergeynikitin wrote on Wed, 06 May 2009 09:28 That is, there is a letter designating the ch or just a couple of characters in the alphabet? I do not understand.

That's the tricky part about it. In Czech alphabet "ch" is single letter, and for example dictionaries are sorted by it, i.e. you have:

banik, cislo, frajer, hroch, chyba, ikra
(notice position of "cislo" and "chyba")

But it is written from regular "c" and "h" letters.

This did lead to a situation where in the digital land the "ch" is always represented by regular "c" and "h" characters, i.e. the single "ch" letter is formed in digital data by regular two characters. This means that in correct cs/sk sort you have to compare not just single character (bytes), but in case you hit "c", you have to look ahead for next character, if it is "h", you have special case.

Unless you use special substitution character for every "c" "h" pair before sort, and then replace it back, you can't cover this behavior by simple order-table.

edit:

If my personal opinion would count, I would discard "ch" immediately from our alphabets, so all the "ch" words would be created from regular "c,h" pairs, i.e. pretty much the current situation except weird sort and weird letter count.

Then again, if my personal opinion would count, there would be no Russian alphabet at all, because I said in school once at Russian lesson: "a nation which is unable to learn+use English 26 letter alphabet will vanish in future". (I was 10 years old at that point and I was so glad ASCII standard existed, so I hope you will understand it a bit)

Since then I admit the utf-8 may be remedy of current status, still if everyone would switch to ASCII 20 years ago and discard all funny chars (and we got much more of them then just "ch", plenty of accented ones too), I bet the current SW would be much smoother ride and lot more cheaper then it is....
