
Subject: New .scd speller file format
Posted by [mirek](#) on Tue, 10 Nov 2009 11:13:50 GMT
[View Forum Message](#) <> [Reply to Message](#)

After working with Koldo on spellers for Basque language, the resulting .scd size was 20MB, which was unacceptable.

Therefore I have developed a new .scd format which has much reduced file size, that Basque is now 5MB, which is not great but completely acceptable.

RichEdit still supports older .scds as well.

Subject: Re: News .scd speller file format
Posted by [koldo](#) on Tue, 10 Nov 2009 15:06:48 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello Mirek

These are good news mainly for so called agglutinative languages like Turkish, Japanese, Georgian and many others like of course Basque, where most words are formed by joining morphemes together.

As you proposed now it is possible to create a place in Google Code or SourceForge where everybody can put and download spelling dictionaries for Upp (mainly with BSD like licenses).

Just to remember that uppsrc/MakeSpellScd package lets to compile a UTF-8 (without BOM) text file with a word per line into a .scd file.

One good thing would be to document this or/and add a small sample about how to use the spelling technology into a program.

Best regards
Koldo

Subject: Re: News .scd speller file format
Posted by [Mindtraveller](#) on Tue, 10 Nov 2009 19:50:31 GMT
[View Forum Message](#) <> [Reply to Message](#)

A little question. May be it will be good idea to make scd file compatible with OpenOffice vocabulary files? So we could use OO community efforts in our projects.

Subject: Re: News .scd speller file format
Posted by [mirek](#) on Tue, 10 Nov 2009 22:13:37 GMT

Mindtraveller wrote on Tue, 10 November 2009 14:50A little question. May be it will be good idea to make scd file compatible with OpenOffice vocabulary files? So we could use OO community efforts in our projects.

You can convert these files to the plain set of words. I even have a tool to do so.

My problem with OO format is that I do NOT really have an idea how to use it for fast spellchecking

Mirek
