

Chybné kódování CSV souboru

Jediné přípustné kódování CSV souboru je UTF-8, což řeší problémy s interoperabilitou na webu, zejména pak s diakritikou a písmeny z různých abeced. Může se ale stát, že váš soubor používá jiné kódování, v českém prostředí zejména Windows-1250 či ISO-8859-2, a proto není validní.

Symptomy

Nejčastějším symptomem je chybné zobrazení diakritiky.



Ověření problému

Chybné kódování lze zjistit buďto použitím [validátoru](#), nebo také otevřením souboru [textovém editoru](#). Na obrázku je kódování ANSI, což odpovídá ISO-8859-2.



Řešení

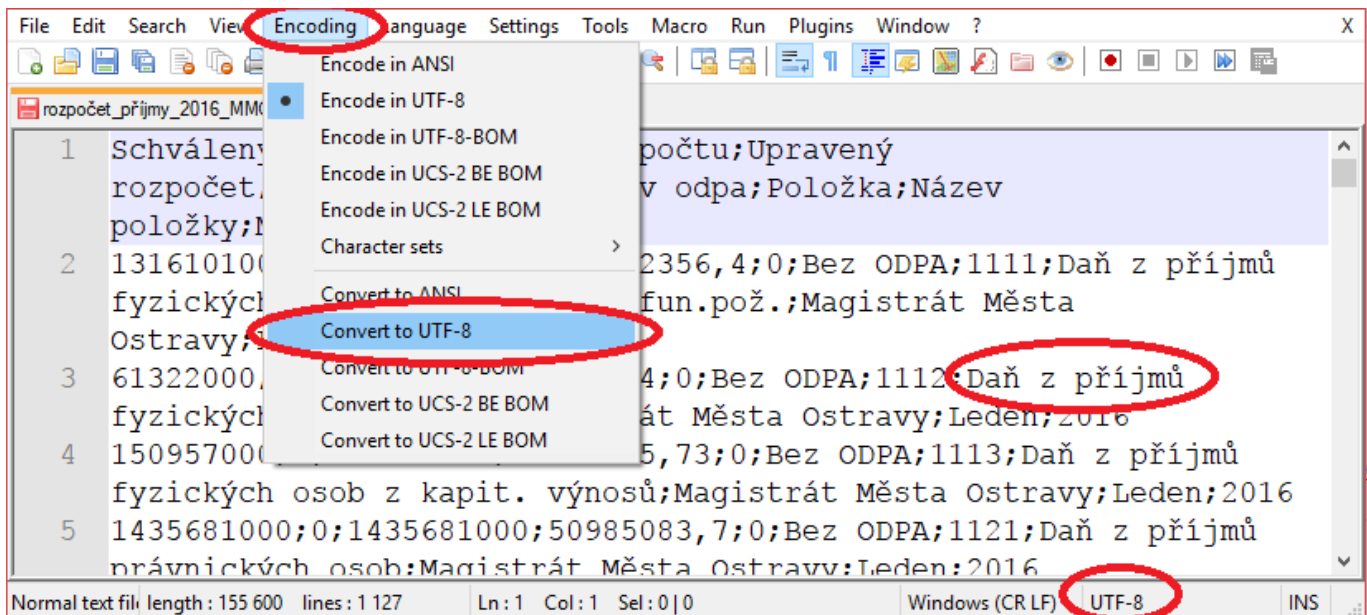
Je třeba zajistit, že soubor je v kódování UTF-8. To lze několika způsoby.

Správné nastavení výstupu u zdroje

Nejjednodušší je zásah u zdroje problému, tedy pokud CSV soubor exportujeme z databáze, nebo generujeme v kódu aplikace, mělo by jít nastavit UTF-8 jako výstupní kódování tam.

Konverze souboru v textovém editoru

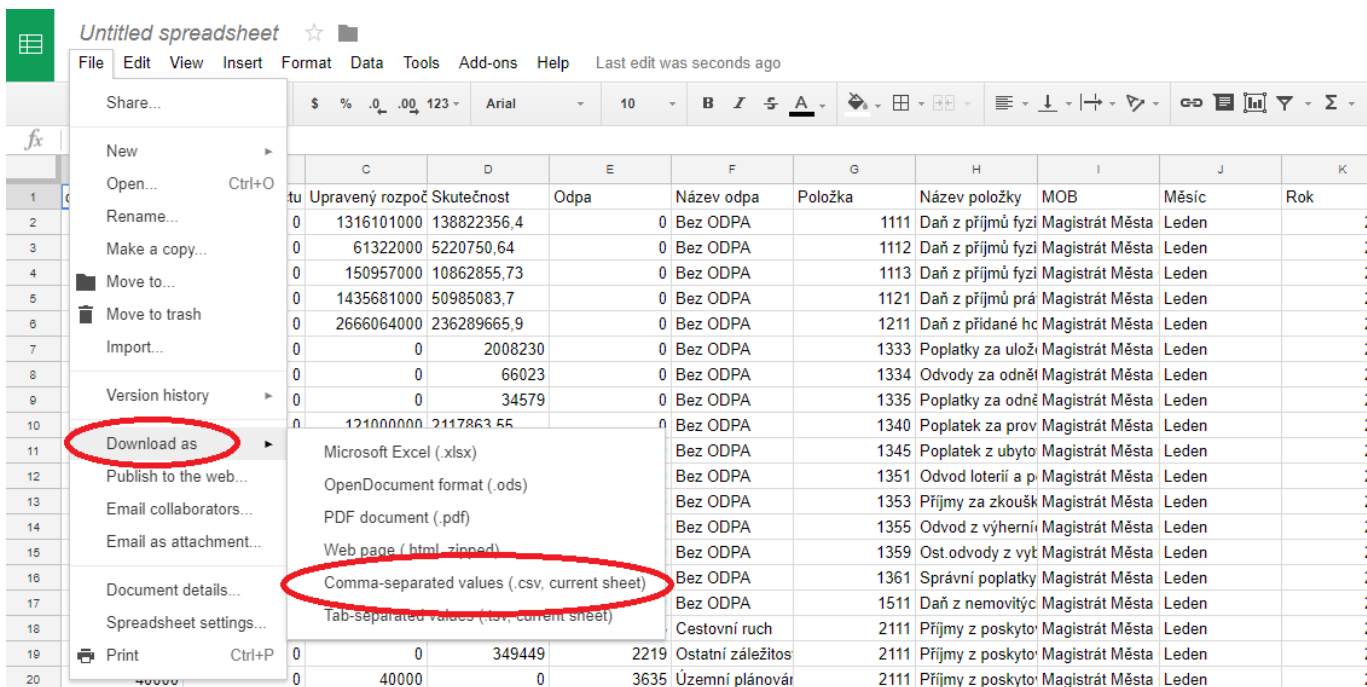
Pokud nemáme přístup ke zdroji, je třeba samotný CSV soubor překódovat, což lze například opět v [textovém editoru](#), jak je vidět na obrázku.



Konverze z pomocí Google Sheets

Můžeme rovněž využít Google Sheets, který si s chybným kódováním poradí a nechá nás stáhnout validní CSV. Budeme potřebovat Google účet, který ale lze zřídit zdarma.

- Do [Google Drive \(Disk Google\)](#) nahrajeme soubor ke konverzi a otevřeme ho pomocí Google Sheets
- Stáhneme data do CSV souboru



From:

<https://opendata.gov.cz/> - **Otevřená data**

Permanent link:

<https://opendata.gov.cz/standardy:csv:chybn%C3%A9-k%C3%B3dov%C3%A1n%C3%AD> 

Last update: **2017/09/04 00:00**