

Strojově nečitelný formát

Strojová čitelnost souboru není dána jeho formátem (CSV, XML, JSON, ...), ale způsobem využití daného formátu. Tedy i zdánlivě strojově čitelný soubor CSV může být ve skutečnosti strojově nečitelný, pokud například vznikl uložením tabulky formátované pro tisk, a tak obsahuje zbytečné prázdné řádky či odsazení.

Podobně je tomu u textových dokumentů. Strojově čitelný textový dokument umožňuje snadný strojový přístup k jednotlivým písmenům textu. Že takový soubor v MS Word může být i úplně strojově nečitelný demonstruje tento

příklad strojově nečitelného textového souboru ve formátu MS Word, který na stránkách místo textu obsahuje obrázky textu jednotlivých stránek, což se strojově zpracovat dá jen těžko, pomocí OCR.

Jiným příkladem je soubor, který má být, ale není, CSV - není to jedna tabulka s fixním počtem sloupců, obsahuje komentáře, prázdné řádky, několik tabulek, a řadu dalších problémů:

```
#OPEN_EXPORT: 2020.4.14.1354
#Datum vytvoření souboru: 15.04.2020 18:52
#Stažením publikovaných dat souhlasíte s podmínkami užití těchto dat.

METADATA
Stanice ID;Jméno stanice;Začátek měření;Konec měření;Zeměpisná
délka;Zeměpisná šířka;Nadmořská výška
P1PKLE01;Praha, Klementinum;01.01.1961;31.12.1993;14,4164;50,0867;190,7
P1PKLE01;Praha, Klementinum;01.01.1994;31.05.2012;14,4164;50,0867;190,7
P1PKLE01;Praha, Klementinum;01.06.2012;31.12.2019;14,416923;50,086634;190,7

PŘÍSTROJE
Přístroj;Začátek měření;Konec měření;Výška přístroje
Teploměr;01.01.1961;31.05.2012;6
Teplotní čidlo;01.06.2012;31.12.2019;6

PRVEK
Průměrná denní teplota vzduchu [T.AVG, °C]

Popis:
Hodnota - hodnota prvku
Příznak - rozšiřující informace o hodnotě

DATA
Rok;Měsíc;Den;Hodnota;Příznak
1961;01;1;,4;
1961;01;2;,3;
1961;01;3;2,2;
```

Last update: 2020/05/22 12:44
špatná-praxe:strojově-nečitelný-formát <https://opendata.gov.cz/%C5%A1patn%C3%A1-praxe:strojov%C4%9B-ne%C4%8Diteln%C3%BD-form%C3%A1t>

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

Permanent link:
<https://opendata.gov.cz/%C5%A1patn%C3%A1-praxe:strojov%C4%9B-ne%C4%8Diteln%C3%BD-form%C3%A1t>

Last update: **2020/05/22 12:44**

