

Educatio Társadalmi Szolgáltató Nonprofit Kft.

AVIR – CSV FORMÁTUM

A CSV állományok formátumának specifikációja

Simon Csaba
2010.08.23.

Dokumentum állapota: Jóváhagyva
Verzió: 1.0

Jelen dokumentum szigorúan bizalmas, tartalma Educatio Társadalmi Szolgáltató Nonprofit Kft. szellemi tulajdonát képezi, felhasználása kizárólag Educatio Társadalmi Szolgáltató Kft. munkatársai számára engedélyezett, akiknek az üzleti kapcsolat értékelésének folyamatában feladatuk van.

TARTALOMJEGYZÉK

1. Formátum.....	3
1.1. Általános.....	3
1.2. Kódolás.....	3
1.3. Adatformátumok	3

1. FORMÁTUM

Az állományok formátumára a következő megkötések érvényesek.

A formátum alapjául a vesszővel tagolt értékeket tárolására használt általános formátum szolgál, melynek leírása itt található: http://en.wikipedia.org/wiki/Comma-separated_values

A dokumentum a leírás által nem vagy nem egyértelműen szabályozott kérdésekben írja le, hogy az AVIR milyen formátumban várja az adatokat.

1.1. Általános

1. Az adatmezők elválasztására használt karakter a vessző.
2. Az adatrekordok (sorok) elválasztására a Windows-os Kocsivissza-soremelés párosítás használandó (CRLF, 0x0D 0x0A). Az egyes adatrekordok nem tartalmazhatnak sortörést.
3. Az utolsó rekord (sor) végén is szerepelnie kell a sortörésnek.
4. A sorok az „id” tartalma oszlop alapján növekvő sorrendben rendezve szerepelnek az állományban.

1.2. Kódolás

1. Az állományok UTF-8 kódolással kell készüljenek.
2. Az állományok elején tilos szerepeltetni a Byte order mark 3 byte-ját (http://en.wikipedia.org/wiki/UTF-8#Byte_order_mark)

1.3. Adatformátumok

1. Számok
 - 1.1. A számok köré nem kell idézőjel.
 - 1.2. Tizedes törtek jelölése: tizedesponttal. Pl. 25.5
2. Dátum és idő
 - 2.1. Dátumot vagy időt tartalmazó cellát idézőjelek közé kell zárni.
 - 2.2. Dátumok jelölése: YYYY-MM-DD (tehát az év 4 számjegyen, hónap két számjegyen és nap két számjegyen, közöttük pedig mínuszjel)
 - 2.3. Időpont jelölése: HH:MM:SS (tehát az óra, perc és másodperc mindegyike két számjegyen ábrázolva, 24 órás módban). Időzóna jelölésére nincs lehetőség.
 - 2.4. Dátum és idő együttes ábrázolása: a dátum és az idő az előzőekben leírt módon, „dátum idő” formában. Pl. "2010-08-23 15:20:00"
3. Szöveges adatok

2010. február 3.

Formátum

- 3.1. Egy szöveges adatot tartalmazó cella tartalma minden esetben idézőjelek között szerepel (függetlenül attól, hogy tartalmaz-e idézőjelet vagy vesszőt).
- 3.2. Egy szöveges cellában a szövegben szereplő idézőjelet meg kell duplázni. Pl. 16, "2010-08-23", "Kovács ""Kokó"" István", 47