

Seznam či strukturovaná hodnota v jedné buňce

Soubor ve formátu CSV by měl být obrazem tabulky v relační databázi, nejlépe tak, aby do ní šel přímo nahrát, a aby se s daty co nejnadhěji pracovalo. Příklad CSV souboru z datové sady „Číselníky pro volby 2016“, který porušuje základní poučku o tom, jak má vypadat databázová tabulka ([první normální forma](#)):

```
VSTRANA,NAZEVCELK,SLOZENI,TYPVS
299,"Koalice CZ, COEX, NEZ, ODA","013,072,088,133",K
300,"Koalice ČSNS, SV SOS","002,101",K
301,"Koalice KSČM, SZ","005,047",K
459,"Sdružení SZ, US-DEU, NK","005,080,102",D
461,"Sdružení NV, NK","080,163",D
462,"Sdružení SNK ED, SD-SN, SOS, SZ, US-DEU,
NK","005,080,102,103,129,143",D
```

Sloupec SLOZENI totiž obsahuje čárkou oddělený seznam ID stran, které tvoří koalici. Pokud bych se tedy chtěl zeptat například na to, které strany tvoří koalici s ID „459“, jsem odkázán na zpracovávání řetězců v databázovém jazyce místo toho, abych se zeptal zcela přirozeně pomocí tabulky vazeb mezi stranami, která by určovala příslušnost do koalice.

Jsou 2 možnosti řešení.

Dělení na 2 datové sady

Toto byly 2 datové sady, „Číselník stran pro volby 2016“ a „Příslušnost stran do koalic pro volby 2016“. CSV druhé datové sady by vypadalo třeba takto:

```
KOALICE,STRANA
299,013
299,072
299,088
299,133
300,002
```

Pro normalizaci databází se používají různě přísné [normální formy](#). [První normální forma](#) zakazuje strukturované hodnoty, tedy například seznamy. Je to proto, že se pak s takovou hodnotou nedá rozumně pracovat, a data se musí před použitím předzpracovávat, rozpadat na více tabulek.

Duplikace řádku

Druhou možností je seznam rozpadnout do jednotlivých řádků, kde hodnoty ve zbylých sloupcích zůstanou stejné. Tomuto postupu se také říká denormalizovaná tabulka.

VSTRANA, NAZEVCELK, SLOZENI, TYPVS

299,"Koalice CZ, COEX, NEZ, ODA","013",K
299,"Koalice CZ, COEX, NEZ, ODA","072",K
299,"Koalice CZ, COEX, NEZ, ODA","088",K
299,"Koalice CZ, COEX, NEZ, ODA","133",K
300,"Koalice ČSNS, SV SOS","002",K
300,"Koalice ČSNS, SV SOS","101",K

From:

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

Permanent link:

<https://opendata.gov.cz/standardy:csv:seznam-v-bu%C5%88ce>



Last update: **2018/11/13 16:46**