D'après sebsauvage.net ♥

# Alors comme ça tu veux faire du CSV ?

Alors comme ça tu veux faire du CSV ? Et tu crois vraiment que tu peux écrire toi-même facilement un parseur de CSV ? Tu crois vraiment qu'il te suffit de faire un bête `split()` sur chaque ligne du fichier pour t'en sortir ??? **T'es vraiment sûr que t'as pensé à tout ?** Il existe de nombreuses espèces de CSV, et certains sont dangereuses et seront sans pitié pour ton code !



| numero | nom | prix | commentaire |
|---|---|---|---|
| 1 | Perceuse | 15 | consomme peu |
| 2 | Marteau | 7 | manche en bois |
| 3 | Pince | 4 | gaînée |

Ok, donc en CSV:

```
numero,nom,prix,commentaire
1,Perceuse,15,consomme peu
2,Marteau,7,manche en bois
3,Pince,4,gaînée
```

Facile, hein ?



Ok, mais alors si le prix est de 7,8, tu fais comment ? Ah oui merde, il faut entourer les cellules par des double-quotes.

| numero | nom | prix | commentaire |
|---|---|---|---|
| 1 | Perceuse | 15 | consomme peu |
| 2 | Marteau | 7,8 | manche en bois |
| 3 | Pince | 4 | gaînée |

```
numero,nom,prix,commentaire
1,Perceuse,15,consomme peu
2,Marteau,"7,8",manche en bois
3,Pince,4,gaînée
```

Fais bien gaffe, tu peux avoir des fichiers CSV avec *toutes* les cellules double-quotée, ou que certaines. C'est pas systématique.



Bien. Et maintenant, si tu veux mettre des double-quote dans une cellule, tu fais comment ? Il faut les doubler:

| numero | nom | prix | commentaire |
|---|---|---|---|
| 1 | Perceuse | 15 | consomme peu |
| 2 | Marteau | 7 | manche en “bois” |
| 3 | Pince | 4 | gaînée |

```
numero,nom,prix,commentaire
1,Perceuse,15,consomme peu
2,Marteau,7,"manche en ""bois"""
3,Pince,4,gaînée
```

Donc une double-quote n'est pas forcément un marqueur de début de cellule. Si tu es à l'*intérieur* d'une cellule ***et*** qu'elle est suivi d'une autre double-quote, alors c'est une simple double-quote à l'intérieur de la cellule. Tu suis toujours ?

Et tu as même le risque que certains enfoirés te fassent ça:

```
numero,nom,prix,commentaire
1,Perceuse,15,consomme peu
2,Marteau,7,manche en ""bois""
3,Pince,4,gaînée
```

Tu me suis toujours ? C'est pas fini.



Et si on veut faire sur plusieurs lignes ?

| numero | nom | prix | commentaire |
|---|---|---|---|
| 1 | Perceuse | 15 | consomme peu |
| 2 | Marteau | 7,8 | manche<br>en<br>“bois” |
| 3 | Pince | 4 | gaînée |

```
numero,nom,prix,commentaire
1,Perceuse,15,consomme peu
2,Marteau,"7,8","manche
en
""bois"""
3,Pince,4,gaînée
```

Et oui, tu peux le dire: Chiotte ! Un retour à la ligne ne veut pas forcément dire “nouvelle ligne

données”. Ça dépend si tu es à l'intérieur d'une chaîne double-quotée ou non.

Histoire de rigoler, n'oublie que certains outils d'export ou tableurs vont te mixer les retours à la ligne `\r\n`, `\r` et `\n` dans un même fichier. Fais attention à la manière dont tu lis et comptes les lignes.



Et avec ça ?

| numero | nom | prix | commentaire |
|---|---|---|---|
| 1 | Perceuse | 15 | consomme peu |
| 2 | Marteau | 7,8 | |
| 3 | Pince | 4 | gaînée |

```
numero,nom,prix,commentaire
1,Perceuse,15,consomme peu
2,Marteau,"7,8"
3,Pince,4,gaînée
```

Si tu fais une boucle en t'attendant à trouver 4 colonnes par ligne, ton programme va directement planter. Certains exporteurs CSV ne mettent pas les cellules vides en fin de ligne et zappent les séparateurs. Oui je sais, c'est chiant.



Et avec ça ?

| numero | nom | prix | commentaire |
|---|---|---|---|
| 1 | Perceuse | 15 | consomme peu |
| 2 | Marteau | 7,8 | |
| 3 | Pince | 4 | gaînée |

```
numero,nom,prix,commentaire
1,Perceuse,15,consomme peu
2,Marteau,"7,8",,,,,,
3,Pince,4,gaînée
```

Oui moi aussi je pense qu'il y a des gens dont l'unique but dans la vie est d'emmerder le monde. Il faudra ignorer les colonnes en trop.



Mais tes problèmes ne sont pas terminés:

| numero | nom | prix | commentaire |
|---|---|---|---|
| 1 | Perceuse | 15 | consomme peu |
| 2 | Marteau | 7,8 | |
| 3 | Pince | 4 | gaînée |

```
1,Perceuse,15,consomme peu
2,Marteau,"7,8"
3,Pince,4,gaînée
```

Mais quel con a omis les noms de colonnes ? Faudra penser à ajouter une option pour prendre en compte (ou non) la première ligne avec les titres des colonnes. C'est pas systématique.



**Tu en veux encore ?**

| numero | nom | prix | commentaire |
|---|---|---|---|
| 1 | Perceuse | 15 | consomme peu |
| 2 | Marteau | 7 | manche en bois |
| 3 | Pince | 4 | gaînée |

```
numero;nom;prix;commentaire
1;Perceuse;15;consomme peu
2;Marteau;7;manche en bois
3;Pince;4;gaînée
```

Mais quel espèce de crétin utilise des **points-virgules** à la place des **virgules** comme séparateur de champ ??? Et oui mon gars, c'est le genre de spécimen que tu rencontres parfois en territoires sauvages. Et tu dois t'apprêter à les combattre: Ton parseur devra avoir un séparateur de colonnes **paramétrable**. S'il est assez (af)fûté, il pourra peut-être même sniffer le séparateur.

Et n'oublie pas qu'il faut toujours double-quoter les chaînes si ton séparateur se retrouve dans la valeur d'un champ.

Et j'ai entendu dire que certaines espèces ont des tabulations comme délimiteurs de champs (TSV), alors sois prudent.



| numero | nom | prix | commentaire |
|---|---|---|---|
| 1 | Perceuse | 15 | consomme peu |
| 2 | Marteau | 7 | manche en bois |
| 3 | Pince | 4 | gaînée |

```
numero,nom,prix,commentaire
1,Perceuse,15,consomme peu
```

```
2,Marteau,7,manche en bois

3,Pince,4,gaînée
```

Oui, des fois il y a des lignes vides. Il faudra les ignorer.



| numero | nom | prix | commentaire |
|---|---|---|---|
| 1 | Perceuse | 15 | consomme peu |
| 2 | Marteau | 7 | manche en bois |
| 3 | Pince | 4 | gaînée |

```
numero,nom,prix,commentaire
1,Perceuse,15,consomme peu
# bla bla bla
2,Marteau,7,manche en bois
3,Pince,4,gaînée
```

Il peut aussi y avoir des commentaires dans les fichiers CSV. Il ne faudra pas importer ces lignes. Et attention si ton parseur indique à quelle ligne il y a une erreur dans le fichier: il faudra compter les lignes de commentaires (et les lignes vides)(et les retours à la ligne à l'intérieur d'une cellule) si tu veux indiquer la bonne ligne.



| numero | nom | prix | commentaire |
|---|---|---|---|
| 1 | Perceuse | 15 | consomme peu |
| 2 | Marteau | 7 | manche en bois |
| 3 | Pince | 4 | gaînée |

```
numero,nom,prix,commentaire
1,Perceuse,15,consomme peu
 # bla bla bla
2,Marteau,7,manche en bois
3,Pince,4,gaînée
```

Alors, commentaire, pas commentaire ? Je sais pas…



Les lignes qui commencent par # ne sont **pas forcément** des commentaires ! Non je ne dis pas ça *exprès* pour t'emmerder. Regarde:

| numero | nom | prix | commentaire |
|---|---|---|---|
| 1 | Perceuse | 15 | consomme peu |

| numero | nom | prix | commentaire |
|---|---|---|---|
| 2 | Marteau | 7,8 | manche<br>#en<br>“bois” |
| 3 | Pince | 4 | gaînée |

```
numero,nom,prix,commentaire
1,Perceuse,15,consomme peu
2,Marteau,"7,8","manche
#en
""bois"""
3,Pince,4,gaînée
```

Attention la ligne #en n'est **PAS** un commentaire, parce que tu es à l'intérieur d'une cellule qui prend plusieurs lignes !

Tu commences à avoir chaud ?

---

# Moralité

**N'ESSAIE PAS DE PARSER LE CSV TOI-MÊME.** Tant que c'est possible, prends une lib solide faite pour ça. (Python a une excellente lib de CSV incluse.)

Et ne te dis pas que ça ne t'arrivera pas, et que toi, tu ne reçois que du CSV bien formaté: Avec la loi de Murphy, s'il y a une possibilité que ça foire, **ça va foirer**.

Si tu veux vraiment le coder toi-même, tu peux s'en sortir en écrivant une machine à état. Mais il faudra bien prendre en comptes tous ces cas. Et peut-être d'autres que j'aurais oubliés.



Vous voulez échangez de grandes quantité de données tabulaires de manière fiable sous forme de fichiers ? **UTILISEZ SQLITE !!!** C'est solide, ultra-rapide, normalisé, multiplatformes, accessible dans tous les langages et remarquablement fiable.

From:
http://www.charpenel.org/wiki/ - **Tutos en vrac**

Permanent link:
**http://www.charpenel.org/wiki/doku.php?id=faire_du_csv_a_la_main&rev=1695202753**

Last update: **2023/09/20 10:39**