

Proposer un contenu

Identifiant

Identifiant

Mot de passe

Mot de passe

☐ Connexion automatique

Se connecter

Pas de compte ? S'inscrire...

Journal : Unicode, NDJSON et bug subtil

Posté par n_e le 19 octobre 2026 à 21:34.

Étiquettes : aucune



Bonjour Nal,

Pendant des années, j'ai utilisé le code suivant pour parser des fichiers NDJSON (un format de fichier qui contient un document JSON par ligne) en node.js :

```
import { stdin } from "node:process";
import readline from "node:readline/promises";

const rl = readline.createInterface(stdin);

for await (const l of rl) {
  const data = JSON.parse(l);
}
```

Cependant il a un bug, saurez-vous le trouver ?

La reproduction

Le bug est exhibé lorsqu'on passe le texte généré ainsi au programme ci-dessus :

```
console.log(JSON.stringify("a\u2028b")); // Af
```

Le bug

[U+2028](#) est un caractère nommé "séparateur de lignes". C'est un caractère qui est autorisé dans les chaînes de caractères JSON, et qui est considéré comme un séparateur de lignes par readline.

Par conséquence le JSON `"a\u2028b"` (où `\u2028` désigne le caractère dont on parle, pas la séquence d'échappement qui le produit) est découpé par readline en `"a` et `b"` qui sont du JSON invalide.

Les caractères U+2028 et U+2029

Unicode a introduit deux caractères, U+2028 (Séparateur de lignes) et U+2029 (Séparateur de paragraphes), que le consortium [a recommandé d'utiliser](#) à la place des caractères issus de ascii (`\n` et `\r\n` selon la plateforme) qui ne distingue pas s'il s'agit d'un retour à la ligne au sein d'un paragraphe ou un nouveau paragraphe.

Aujourd'hui ces caractères ne semblent pas gérés : dans vim et dans le terminal ils sont affichés comme des espaces, et Visual Studio Code produit l'avertissement suivant :

```
The file 'u2028.txt' contains one or more unusual
line terminator characters, like Line Separator (LS) or
Paragraph Separator (PS).

It is recommended to remove them from the file.
```

Les specs

Concernant les caractères autorisés dans une chaîne de caractères JSON, tout est autorisé sauf les caractères de contrôle ascii (dont notamment `\n`), le caractère `\` et les guillemets fermants, qui doivent être échappés.

Concernant le readline de Node.js, le comportement n'est [pas documenté](#) (la doc ne mentionne que `\n` et `\r\n`), mais le problème a [été introduit](#) dans Node 24.0.0, malheureusement sans explication, et alors que la tendance semblait davantage être à ne plus utiliser ces caractères.

(1 commentaire).

Markdown

EPUB

ASCII's not dead ?

Posté par Marotte ✨ le 20 octobre 2026 à 04:12.
Évalué à 3 (+0/-0).

Les caractères de contrôle ASCII sont interdit dans JSON ? Je pense aux caractères `\x1c` `\x1d` `\x1e` et `\x1f`

Le programme `unicode` me sort un truc assez surprenant, il me les désigne tous les quatre comme des “paragraph separators” ou “segment separator” pour `\x1f` ou encore “information separators” four three two one...

```
$ unicode -v -x 1c 1d 1e 1f | grep -f
U+001C INFORMATION SEPARATOR FOUR
Bidi: B (Paragraph Separator)
U+001D INFORMATION SEPARATOR THREE
Bidi: B (Paragraph Separator)
U+001E INFORMATION SEPARATOR TWO
Bidi: B (Paragraph Separator)
```

U+001F INFORMATION SEPARATOR ONE
Bidi: S (Segment Separator)

À ma connaissance il s'agit de, respectivement, Field Separator, Group Separator, Record Separator et Unit Separator. Ces caractères pourraient-ils résoudre ton problème ?

Répondre

Envoyer un commentaire

Suivre le flux des commentaires

Note : les commentaires appartiennent à celles et ceux qui les ont postés. Nous n'en sommes pas responsables.

Revenir en haut de page

Derniers commentaires

- ASCII's not dead ?
- Opération au guiche...
- Re: Mmmh
- DoH système vers A...
- Re: Peut-être plus si...
- Re: Sécurisation
- Re: Solution de facilité
- Re: Sécurisation
- Re: Sécurisation
- Re: C'est un bon début
- Re: Ces applications...
- Re: Forks

Étiquettes (tags) populaires

- intelligence_artificielle
- slop
- merdification
- états-unis
- grands_modèles_de...
- administration_fran...
- bigtech
- capitalisme
- spam
- python
- chine
- politique

Sites amis

- Agenda du Libre
- April
- Éditions D-BookeR
- Éditions Diamond
- Éditions Eyrolles
- Éditions ENI
- En Vente Libre
- Framasoft
- La Quadrature du Net
- Lea-Linux
- Open Source Initiative
- Imprimerie Grafik Plus

À propos de LinuxFr.org

- Mentions légales
- Faire un don
- L'équipe de LinuxFr...
- Informations sur le s...
- Aide / Foire aux que...
- Suivi des suggestion...
- Wiki du site
- Règles de modération
- Statistiques
- API pour le développ...
- Code source du site
- Plan du site