

I created a post where the excerpt was copy&pasted from a pdf document.

When pasting the text, the "fi" glyph disappears (e.g. "specification" is copied over as "specication", this is a common problem, see for instance: <http://superuser.com/questions/375449/why-does-the-text-fi-get-cut-when-i-copy-from-a-pdf-or-print-a-document>).

To be more precise, the "fi" glyph is replaced with the .notdef glyph. The .notdef glyph is not visible in the Edit Post screen nor when viewing the post but it is stored in the database (rendered as a white square, the most common representation for this glyph).

The problem is that, while the glyph is properly filtered when viewing the post, it is not when creating the RSS feed so it breaks it.

For instance, when trying to access it with Google Chrome I get: This page contains the following errors:

```
error on line 29 at column 25: Input is not proper UTF-8, indicate encoding !
```

```
Bytes: 0x0C 0x66 0x69 0x63
```

I've been able to reproduce the problem on several sites.