

# Protege and UTF-8

## Protégé and UTF-8

Protégé is utf-8 compatible, which means it can process and display utf-8 characters. In the course of editing, many editors prepare their concepts or concept information in Microsoft Word or Excel. They then copy from Microsoft and paste into Protégé. This can cause problems because Microsoft is not purely utf-8 compatible. The paste operation can introduce characters that Protégé does not know how to process. The instructions below show how to avoid these problems.

### Table of characters in Microsoft and their utf-8 equivalents

Thanks to [Interwingly](#) for providing this table.

If you are running on a Microsoft platform, or cut and paste from documents produced by Microsoft software, or even allow comments to be posted by people who might be doing one of the above, you need to be aware of the differences, summarized by the following table.

| character | win-1252 decimal | win-1252 hex | win-1252 octal | unicode html | unicode xml | unicode url |
|-----------|------------------|--------------|----------------|--------------|-------------|-------------|
| €         | 128              | 80           | 200            | &euro;       | &#x20AC;    | %E2%82%AC   |
| ,         | 130              | 82           | 202            | &sbquo;      | &#x201A;    | %E2%80%9A   |
| f         | 131              | 83           | 203            | &fnof;       | &#x192;     | %C6%92      |
| „         | 132              | 84           | 204            | &bdquo;      | &#x201E;    | %E2%80%9E   |
| ...       | 133              | 85           | 205            | &hellip;     | &#x2026;    | %E2%80%A6   |
| †         | 134              | 86           | 206            | &dagger;     | &#x2020;    | %E2%80%A0   |
| ‡         | 135              | 87           | 207            | &Dagger;     | &#x2021;    | %E2%80%A1   |
| ^         | 136              | 88           | 210            | &circ;       | &#x2C6;     | %CB%86      |
| ‰         | 137              | 89           | 211            | &permil;     | &#x2030;    | %E2%80%B0   |
| Š         | 138              | 8A           | 212            | &Scaron;     | &#x160;     | %C5%A0      |
| ‹         | 139              | 8B           | 213            | &lsaquo;     | &#x2039;    | %E2%80%B9   |

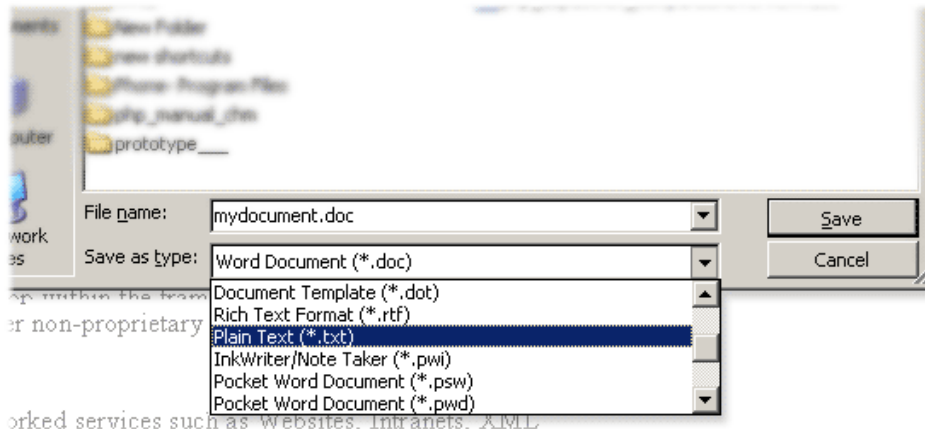
|   |     |    |     |          |          |           |
|---|-----|----|-----|----------|----------|-----------|
| Œ | 140 | 8C | 214 | &OElig;  | &#x152;  | %C5%92    |
| Ž | 142 | 8E | 216 | &x17D;   | &#x17D;  | %C5%BD    |
| ‘ | 145 | 91 | 221 | &lsquo;  | &#x2018; | %E2%80%98 |
| ’ | 146 | 92 | 222 | &rsquo;  | &#x2019; | %E2%80%99 |
| “ | 147 | 93 | 223 | &ldquo;  | &#x201C; | %E2%80%9C |
| ” | 148 | 94 | 224 | &rdquo;  | &#x201D; | %E2%80%9D |
| • | 149 | 95 | 225 | &bull;   | &#x2022; | %E2%80%A2 |
| – | 150 | 96 | 226 | &ndash;  | &#x2013; | %E2%80%93 |
| — | 151 | 97 | 227 | &mdash;  | &#x2014; | %E2%80%94 |
| ˜ | 152 | 98 | 230 | &tilde;  | &#x2DC;  | %CB%9C    |
| ™ | 153 | 99 | 231 | &trade;  | &#x2122; | %E2%84%A2 |
| š | 154 | 9A | 232 | &scaron; | &#x161;  | %C5%A1    |
| › | 155 | 9B | 233 | &rsaquo; | &#x203A; | %E2%80%BA |
| œ | 156 | 9C | 234 | &oeelig; | &#x153;  | %C5%93    |
| ž | 158 | 9E | 236 | &#x17E;  | &#x17E;  | %C5%BE    |
| ÿ | 159 | 9F | 237 | &Yuml;   | &#x178;  | %C5%B8    |

|   |      |      |  |         |  |        |
|---|------|------|--|---------|--|--------|
|   | 160  | A0   |  | &nbsp;  |  | %C2%A0 |
| – | 8211 | 2013 |  | &#8211; |  |        |
| — | 8212 | 2014 |  | &#8212; |  |        |
| ' | 8217 | 2019 |  | &#8217; |  |        |
|   |      |      |  |         |  |        |

## Converting Microsoft Characters to UTF-8 in Word 2003

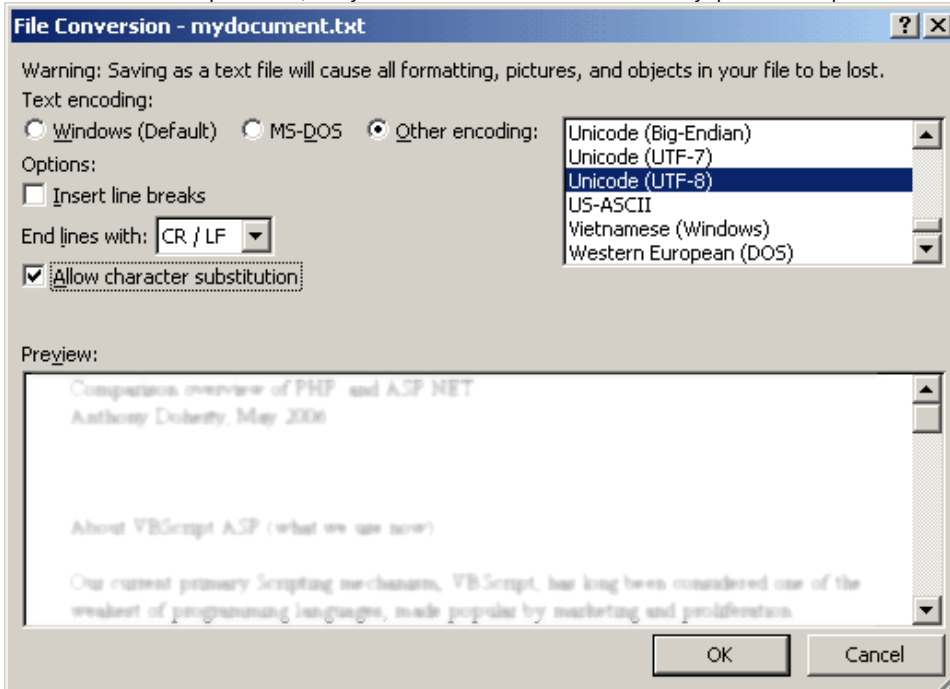
Thanks to [Liverpool John Mores University](#) for providing the following instructions:

- Once your editing in Word is complete, choose **File->Save As...**



- Choose from the format drop-menu the option '**Plain Text (.txt)\***'
- Save the file to a known location, your desktop for example.
- Before the file saves, a dialog box will appear asking you about encoding: Choose '**other encoding**'.
- Then make sure you check the '**Allow Character Substitution**' box.

- Your document is then previewed, and you will see all characters such as 'curly quotes' are replaced with 'safe' ones.

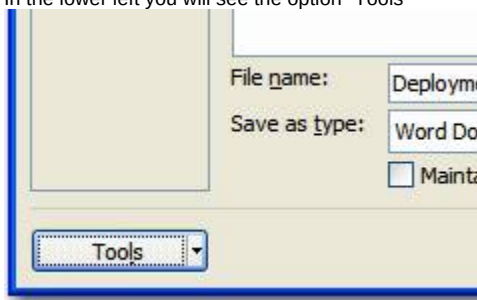


You can then open the saved .txt file and safely copy the contents you require into a web page that uses UTF-8 encoding.

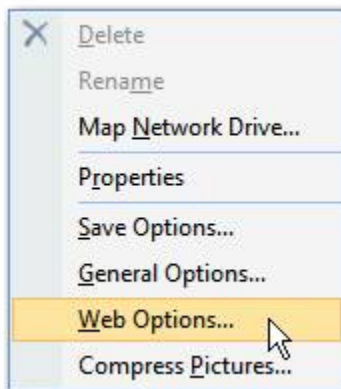
## Converting Microsoft Characters to UTF-8 in Word and Excel 2007

For Word and Excel 2007, the instructions are the same

- Go to File > Save As
- In the lower left you will see the option "Tools"



- Within the Tools drop down, select Web Options



- In the Web Options dialog, go to the Encoding tab and select UTF-8

