

Sometimes my HTML output is missing some text that is in Chinese, Japanese or Korean (CJK) fonts. What can cause this and how can it be fixed?

Some fonts in PDF files use predefined mappings between character encodings and specific, predefined character identifier sets. These mappings are known as character maps (CMaps) and are a standard part of the PDF specification. Because of their large size, these character maps are stored in files that are provided with Acrobat and other Adobe Systems products (and PDF Alchemist), and are usually referenced by name from PDF files when needed.

PDF Alchemist needs to be supplied a folder with these CMap files in order to properly resolve these references. If this folder is not supplied, these references cannot be resolved and the text using these CMaps will not be able to be mapped to Unicode. The characters will be missing from the HTML output file.

These CMap files are contained within the `cmaps` folder in your PDF Alchemist package. Use the `-cmap` commandline parameter or the `cmapDir` API parameter to supply these to PDF Alchemist. A trailing slash is required for this directory, and we recommend an absolute path for most reliable results, like this:

```
C:\DataLogics\PDFAlchemist\x64\cmaps\
```

So the command might look like this:

```
C:\DataLogics\PDFAlchemist>pdfalchemist "Test.PDF" C:\DataLogics\PDFAlchemist\Export -cmap C:\DataLogics\PDFAlchemist\x64\cmaps\
```

In this example you are using the product to convert a file called `Test.PDF` to HTML, and telling PDF Alchemist to look for cmap files in the directory provided with the installation package.