

Article

[Eduard Lebedyuk](#) · Feb 2, 2017 3m read

Converting documents with Caché and LibreOffice

In this article I'll show you how you can easily convert between many different document formats, these to be precise:

- in: html, odt, doc, docx, rtf, sdw, txt...
- out: html, odt, docx, rtf, pdf, txt, tex, xhtml...

and many more, check out this [table](#). We'll use [LibreOffice](#) as a conversion engine.

LibreOffice Installation

On windows go to the [download](#) page, application installs like pretty much any windows takes 200 mb.

On linux either use package manager (recommended) or go to the [download](#) page

- Ubuntu: apt-get install libreoffice-core libreoffice-write
- RHEL: yum install libreoffice-core libreoffice...

Note that on linux you'll need at least version 4 (current is 5.2, if you're on old distro, refer to this [guide](#)).

Post installation

After installation is done, make sure that Caché can access "soffice" application

- On windows add "C: / Program Files (x86) / LibreOffice 5 / program" (may differ due to version installed) to system PATH, if you're running Caché with default settings, or to the user PATH under which you run Caché if you modified that setting. [Guide](#).
- If you installed via package manager, "soffice" should be in path, otherwise add to PATH manually

Use

[Import code](#) into Caché and you're ready to `gconvert`.

Call from the terminal:

```
set sc = ##class(Converter.LibreOffice).convert(source, target, format)
write $System.Status.GetErrorText(sc)
```

Where:

- source - file to convert

- target - result file
- format - specification for target file. Possible values: docx,html,mediawiki, csv,pptx,ppt,wmf,emf,svg,xlsx,xls. More possible values [here](#).

Example

To convert doc into docx call:

```
set sc = ##class(Converter.LibreOffice).convert("C:\temp\1.doc", "C:\temp\1.docx", "docx")
write sc
>1
```

Code

Here's some interesting snippets:

```
/// Get path to libreoffice/soffice
ClassMethod getSO()
{
    if $$$isWINDOWS {
        set path = "soffice"
    } else {
        set path = "export HOME=/tmp && unset LD_LIBRARY_PATH && soffice"
    }
    return path
}
```

Note, that in linux we additionally set HOME variable (it should be writable to Caché user) and remove LD_LIBRARY_PATH variable. This is important for LibreOffice (particularly in web context). If you need LD_LIBRARY_PATH variable, add additional call to reset it later.

```
/// Convert a file %1 into %2 format and place it into %3 directory (name equal to %1
name, extension = %2) using %4 - LibreOffice
Parameter COMMAND = "%4 --headless --writer --convert-to %2 --outdir %3 %1";

/// Convert source into format and place it into targetDir
ClassMethod executeConvert(source, targetDir, format) As %Status
{
    // Libreoffice needs targetDir without last slash
    set:$e(targetDir,*)=..#SLASH targetDir = $e(targetDir, 1, *-1)

    set timeout = 100
    set cmd = $$$FormatText(..#COMMAND, source, format, targetDir, ..getSO())
    return ..execute(cmd, timeout)
}
```

\$\$\$FormatText macro can be used to make OS commands more readable (use it instead of concatenation!).

Conclusions

If you need to convert between a large number of document formats this project can be useful.

Links

- [Code on GitHub](#)

[#Best Practices](#) [#Caché](#) [#Tools](#)

Source URL: <https://community.intersystems.com/post/converting-documents-cach%C3%A9-and-libreoffice>