

Using fontspec package with tex4ht and LuaLaTeX

Michal Hoftich

August 1, 2015

Contents

1	Introduction	1
2	Using alternative4ht	2
3	Limitations	4

You can download this guide in PDF/TeX.

1 Introduction

The fontspec package provides support for OpenType fonts in unicode LaTeX formats, Xe_{La}TeX and Lua_{La}TeX. This enables full support for various non-latin scripts, such as Indic, CJK or Arabic. Huge problem is that it is not supported by tex4ht, convertor from _{La}TeX to HTML.

The reason is that conversion is done by parsing special DVI file with HTML instructions. It is done by tex4ht command¹. This command unfortunately doesn't support OpenType fonts, failing to do any conversion at all. This needs to be fixed in the source code, but as no person with knowledge of C language, DVI format internals and understanding of literate sources of tex4ht is known to exist, it is unlikely that this happens in foreseeable future.

Some possible solutions which doesn't need to fix tex4ht exists:

1. For documents which use only scripts supported by standard pdf_{La}TeX, one can modify the document and use right arguments for inputenc and fontenc packages:

```
\ifdefined\HCode% detect tex4ht
\usepackage[utf8]{inputenc}
```

¹It is little bit confusing that name *tex4ht* means two things: whole system including _{La}TeX packages and some command line applications, tex4ht is also one of these command line applications. The command name will be printed in monospace font in this text to avoid confusion

```

\usepackage[T1]{fontenc}
\else
\usepackage{fontspec}
\setmainfont{TeX Gyre Termes}
\fi

```

this method works best for European languages

2. Use lua4ht, which is an experimental replacement for tex4ht written in Lua. It doesn't support all tex4ht features though. See an example at [TeX.sx](#).
3. Use alternative version of fontspec and other packages provided by helpers4ht. This method will be described later in this document.

2 Using alternative4ht

The alternative4ht package is included in helpers4ht, see the linked page for installation instructions. It provides means to support packages which causes tex4ht to fail. Usually these packages directly include PDF instructions and it confuses tex4ht. In order to minimize a need to modify a document, it enables to load alternative version of a problematic package with tex4ht. This alternative version usually provides main commands which may be used in a document, but the implementation is totally different, with focus on tex4ht support.

```

\usepackage{alternative4ht}
\altusepackage{fontspec}
\setmainfont{TeX Gyre Termes}
\newfontfamily\greekfont{Linux Libertine O}
\newfontfamily\russianfont{Linux Libertine O}
\newfontfamily\hindifont{Siddhanta}
\altusepackage{polyglossia}
\setmainlanguage{english}
\setotherlanguage{czech}
\setotherlanguage{greek}
\setotherlanguage{russian}
\setotherlanguage{hindi}
....
\begin{document}

\begin{czech}
  Czech text
\end{czech}

\textgreek{Greek text}

\begin{russian}

```

```

Russian text
\end{russian}

\begin{hindi}
Hindi text
\end{hindi}

```

I haven't included actual text in the listings, as it would be too long. The selected text is first paragraph from Wikipedia pages about Prague in various languages. As you can see, Polyglossia package provides two ways for including text in other language: `\text<lang>` or `\begin{<lang>}`. The first method is better for smaller passages of text, possibly within paragraph, the latter is good for longer passages of text. Note that it always starts a new paragraph. We also need to provide a font family for each language as is shown with all these `\newfontfamily\<lang>font` commands.

A result with some real world texts:

Praha je hlavní a současně největší město České republiky a 15. největší město Evropské unie. Leží mírně na sever od středu Čech na řece Vltavě, uvnitř Středočeského kraje, jehož je správním centrem, ale jako samostatný kraj není jeho součástí. Je sídlem velké části státních institucí a množství dalších organizací a firem. Sídlí zde prezident republiky, parlament, vláda, ústřední státní orgány a jeden ze dvou vrchních soudů. Mimoto je Praha sídlem řady dalších úřadů, jak ústředních, tak i územních samosprávných celků; sídlí zde též ústředí většiny politických stran a centrály téměř všech církví, náboženských a dalších sdružení s celorepublikovou působností registrovaných v ČR.

Η Πράγα (τσέχικα: Praha), είναι η πρωτεύουσα και μεγαλύτερη πόλη της Τσεχίας. Χτισμένη στον ποταμό Μολδάβα (Vltava), στην κεντρική Βοημία, έχει 1,2 εκατομμύριο κατοίκους. Αποκαλείται επίσης «η χρυσή πόλη» και «μητέρα των πόλεων». Από το 1992, το ιστορικό κέντρο της Πράγας ανήκει στον κατάλογο μνημείων παγκόσμιας κληρονομιάς της UNESCO.

Прага (чеш. Praha [ˈpraɦa]) — город и столица Чехии; административный центр Среднечешского края и двух его районов — Прага-Восток и Прага-Запад. Образует самостоятельную административную единицу страны.

प्राग यूरोप के चेकोस्लोवाकिया देश की राजधानी है।

Compile the document with

```
make4ht -ul filename
```

If your browser supports hyphenation, you can see that some texts are hyphenated. Not all languages are supported by all browsers, unfortunately.

3 Limitations

This method works only with Lua^LTeX at the moment. It uses a node callback to replace all characters with HTML entities, which are later translated back to unicode values by `tex4ht`. I don't know about any easy solution for Xe^LTeX, other than providing huge file with `\DeclareUnicodeCharacter` for all possible characters. This doesn't sound right to me, so I am open to any more realistic ideas.

I provided definitions only for most basic commands, if you find any undefined commands in your documents, please let me know and I will try to provide reasonable definition for them².

²You can provide definitions as well, of course :)