

Taming UTF-8 in pdfTeX

Frank Mittelbach

Abstract

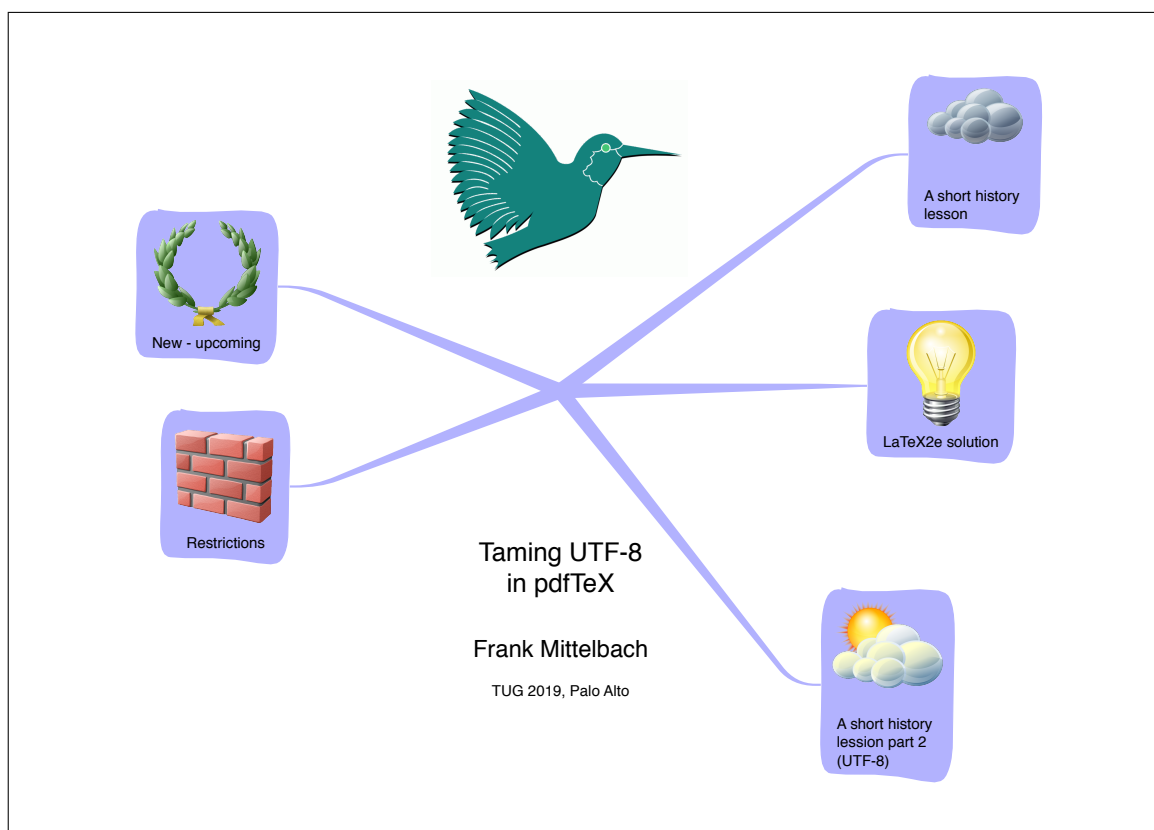
To understand the concepts in pdfLaTeX for processing UTF-8 encoded files it is helpful to understand the models used by the TeX engine and earlier attempts used by LaTeX on top of TeX. The talk provides a short historical review of that area and explains

- how it is possible in a TeX system that only understands 8-bit input to nevertheless interpret and process UTF-8 files successfully;
- what the obstacles are and how they can be overcome and
- what restriction will remain if one doesn't switch to a Unicode-aware engine such as LuaTeX or XeTeX.

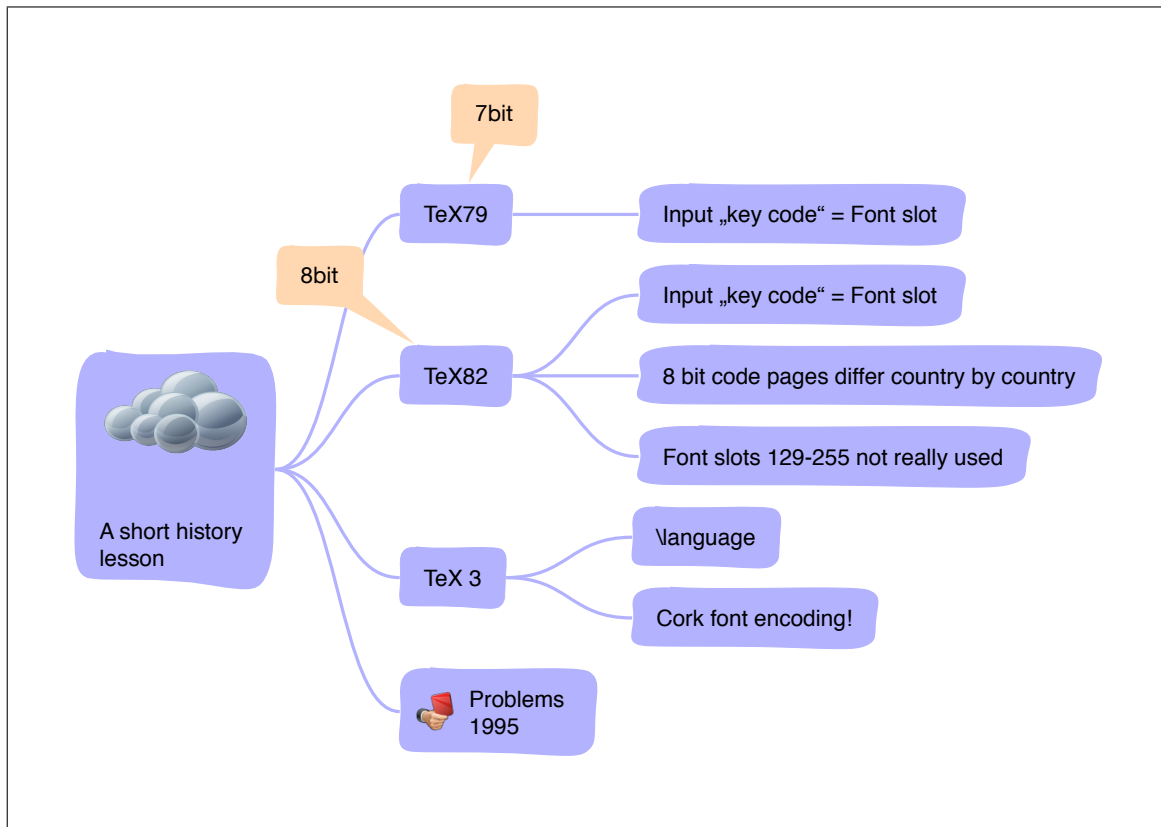
It will finish with an overview about the improvements with respect to UTF-8 handling that will be activated in LaTeX within 2019 and explains how they can already be tested today.

The slides have been retrospectively constructed from the mindmap used during the presentation.

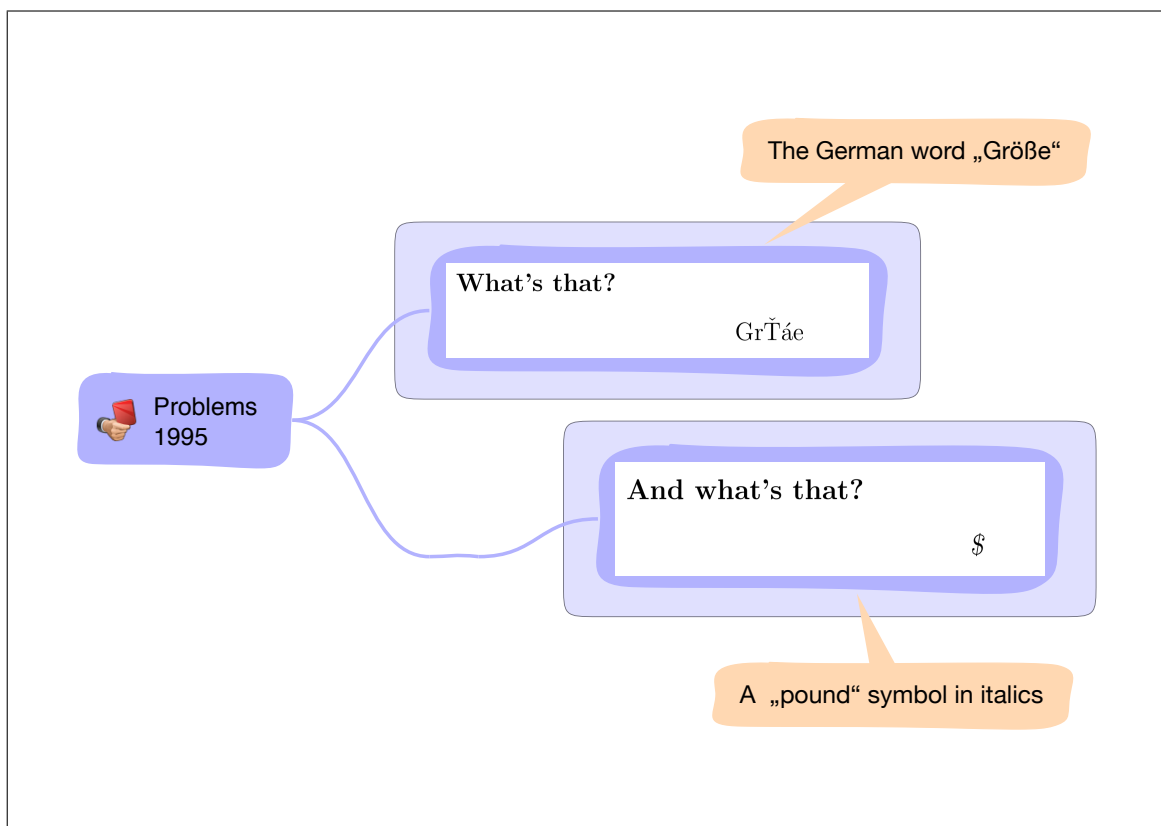
◇ Frank Mittelbach
<https://www.latex-project.org>



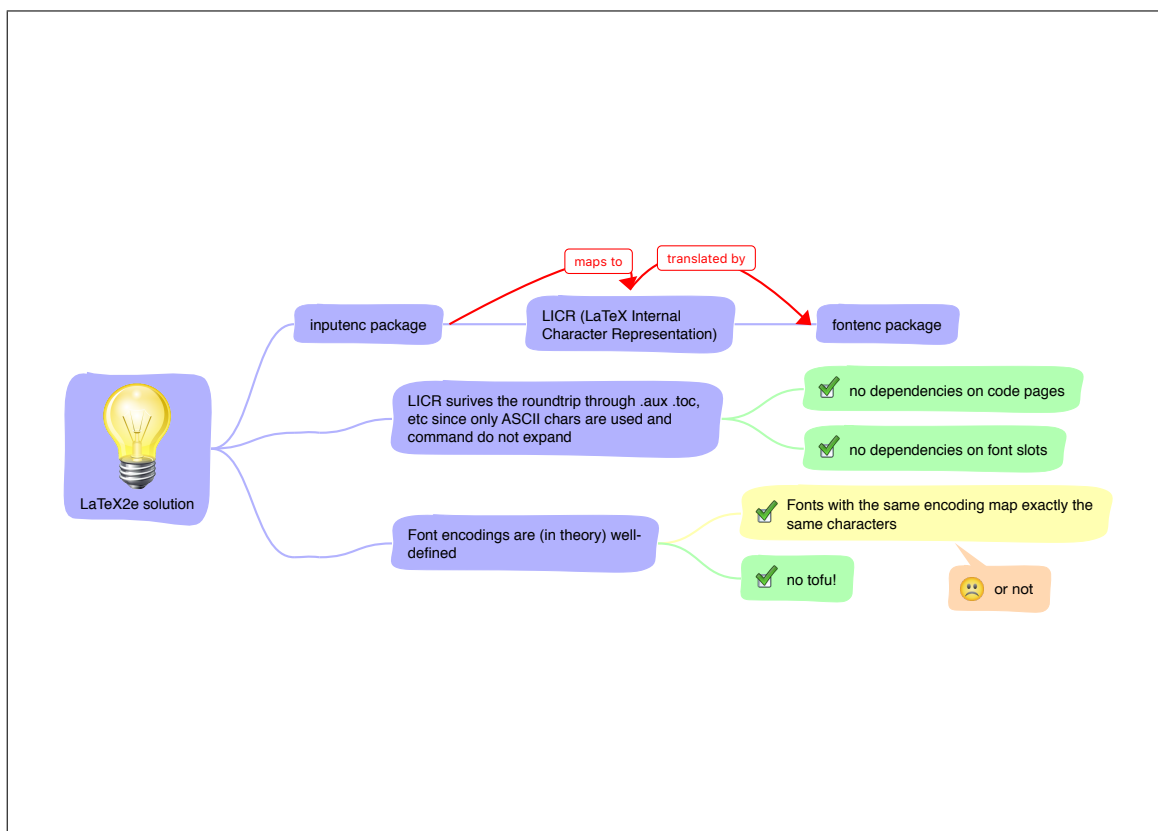
Slide #1



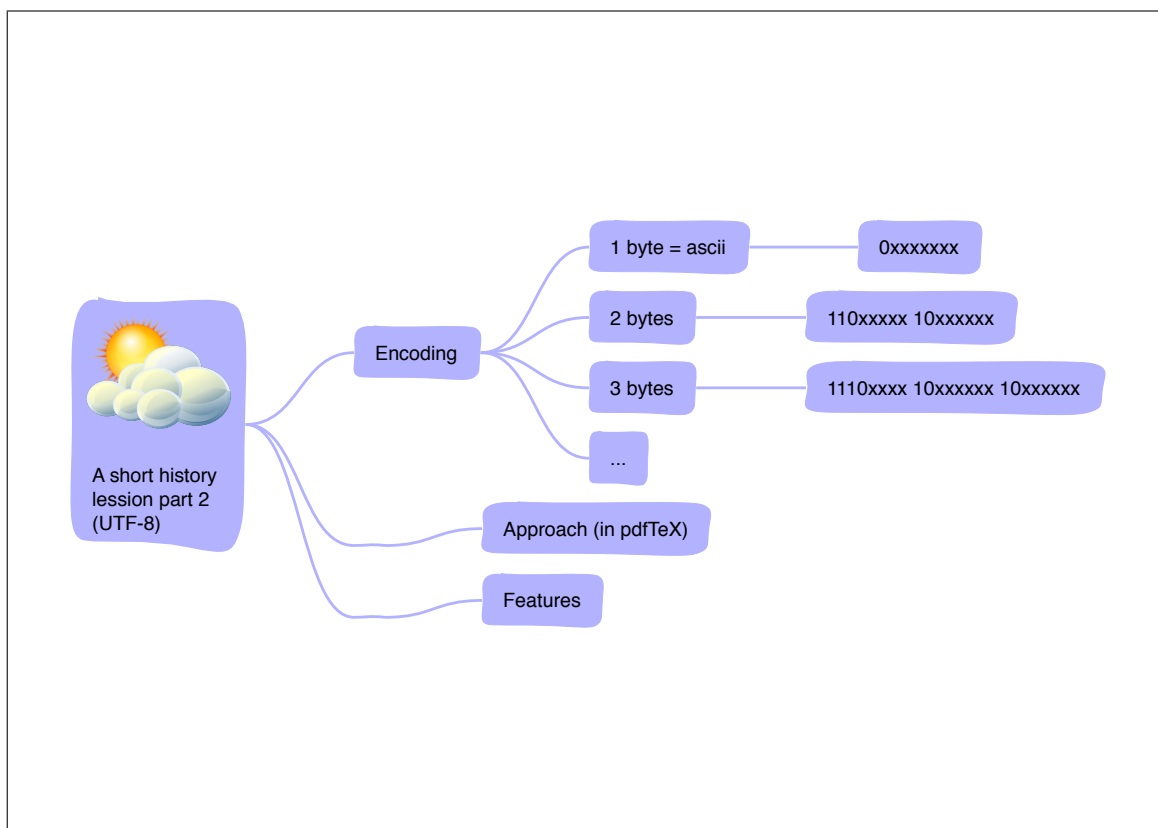
Slide #2



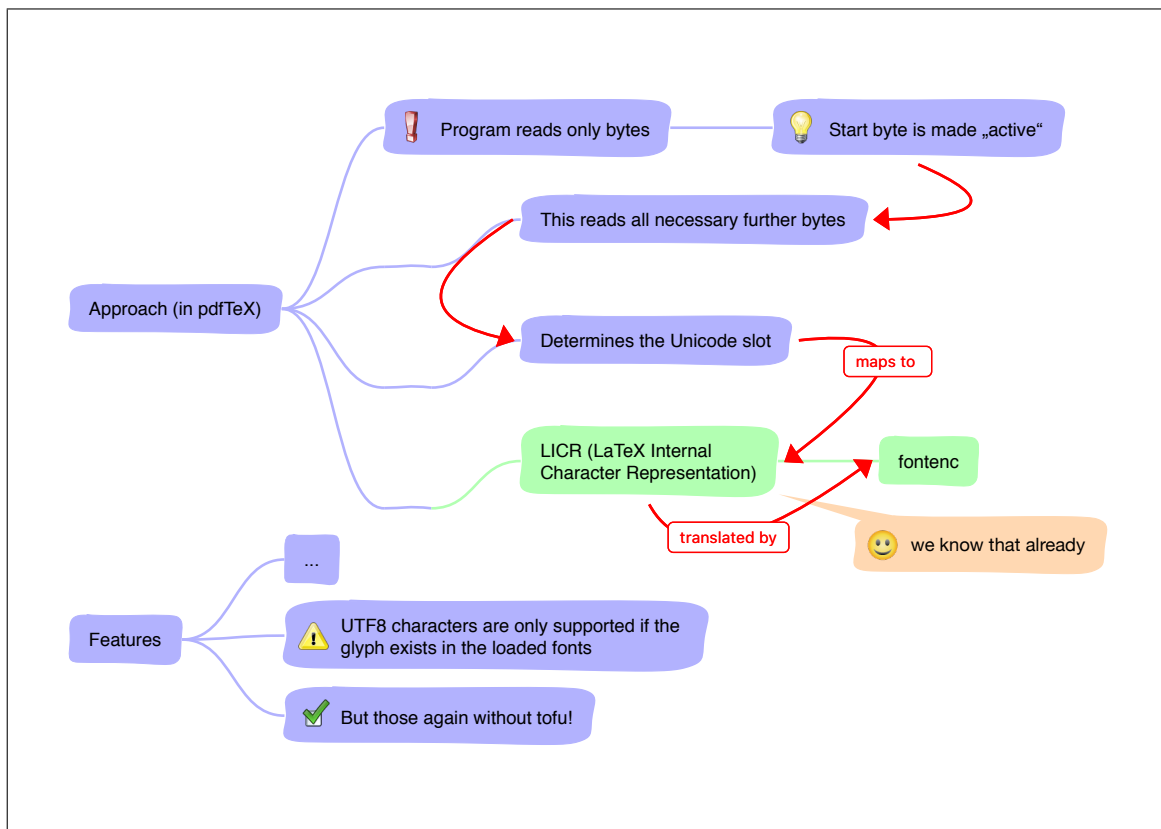
Slide #3



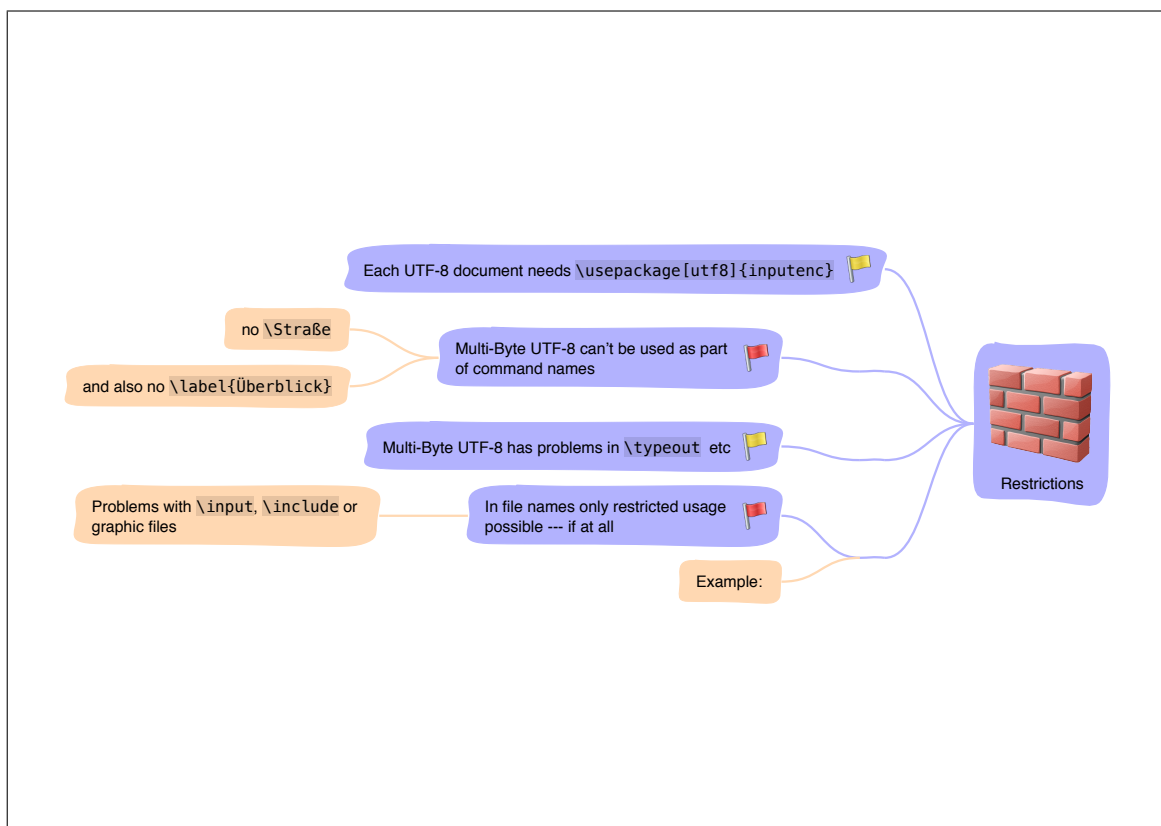
Slide #4



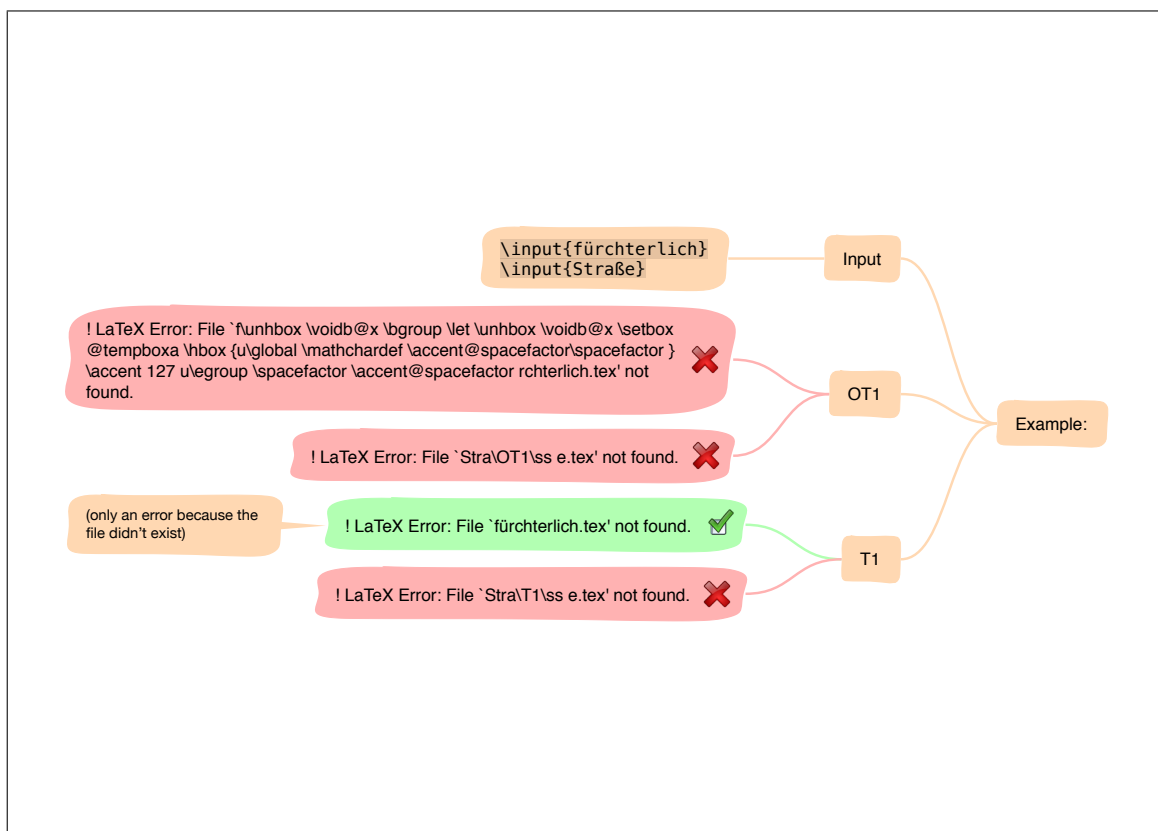
Slide #5



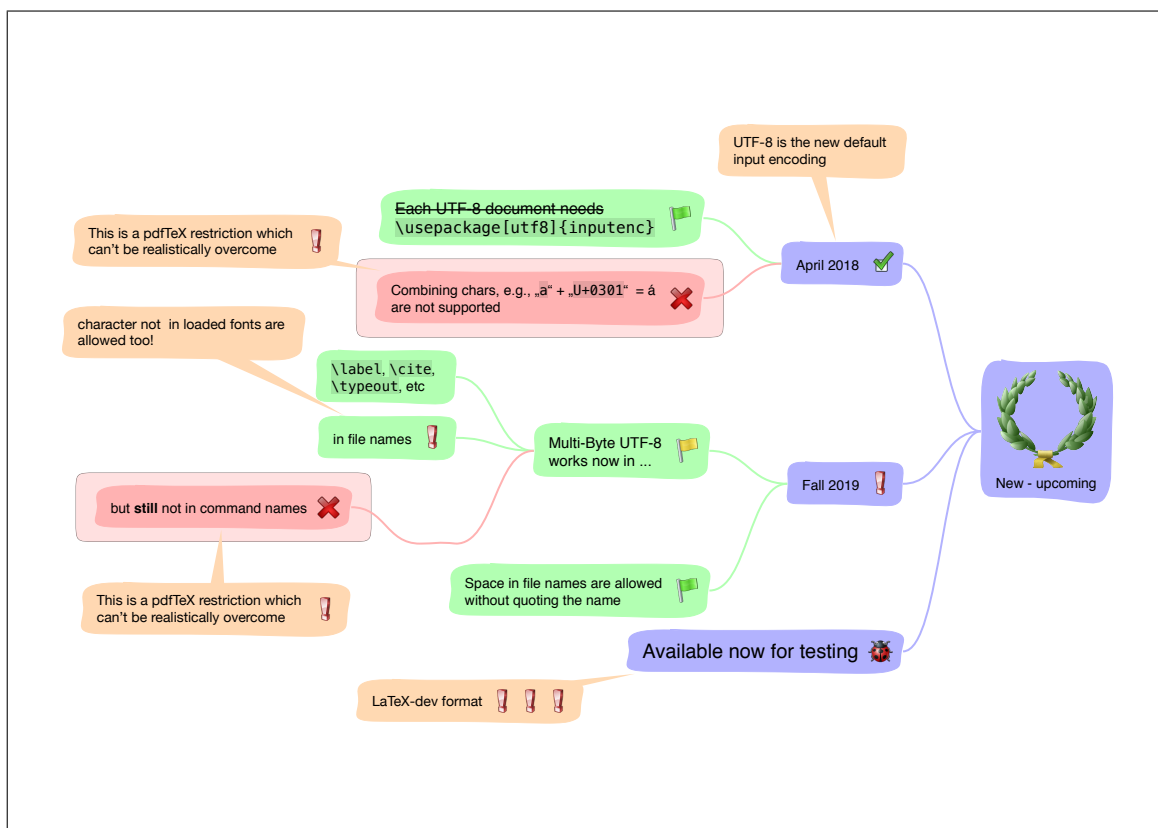
Slide #6



Slide #7



Slide #8



Slide #9