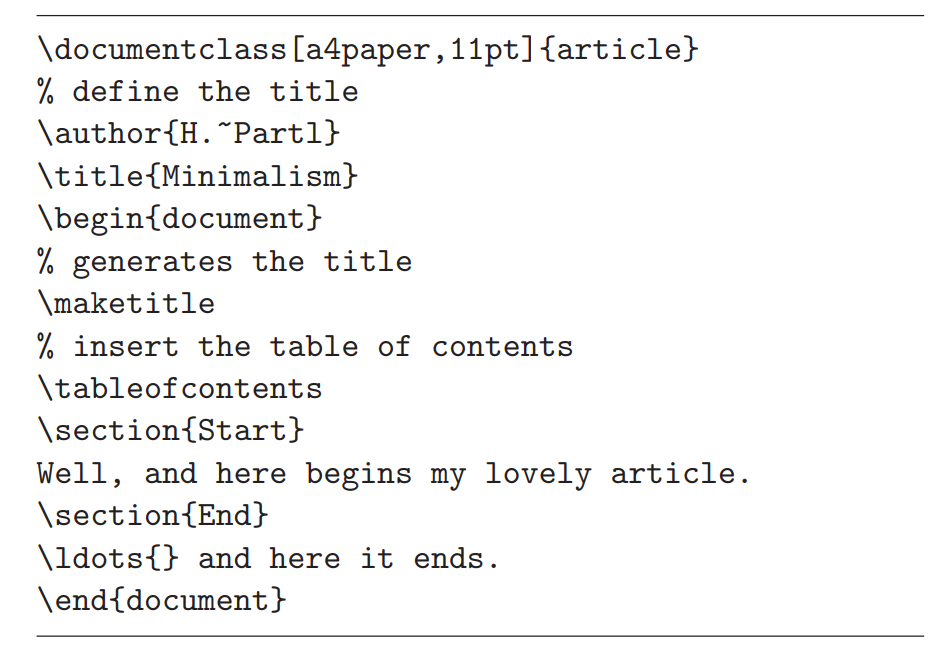
**实验2：基于正则表达式的LaTex编译器**

1. **预备知识**
2. **Tex**

TeX是由著名的计算机科学家Donald E. Knuth（高德纳）发明的排版系统，利用TeX可以很容易地生成高质量的dvi文件，打印输出。利用dvips，dvipdfmx，pdfLaTeX等程序生成pdf，ps文件，LaTeX2html生成html文件。 它在学术界十分流行，特别是数学、物理学、统计学与计算机科学界。TeX被普遍认为是一个很好的 排版工具，特别是在处理复杂的数学公式时。利用诸如是LaTeX等终端软件，TeX就能够排版出精美的文本。

1. **LaTex**

LaTeX是一种基于ΤeΧ的排版系统，由美国计算机学家莱斯利·兰伯特（Leslie Lamport）在20世纪80年代初期开发，利用这种格式，即使使用者没有排版和程序设计的知识也可以充分发挥由TeX所提供的强大功能，能在几天、甚至几小时内生成很多具有书籍质量的印刷品。对于生成复杂表格和数学公式，这一点表现得尤为突出。因此它非常适用于生成高印刷质量的科技和数学类文档。这个系统同样适用于生成从简单的信件到完整书籍的所有其他种类的文档。



一个典型的LaTex源文件

**CTex（中文LaTex发行版）安装网址**：http://www.ctex.org/HomePage

1. **网页到PDF转换工具pyFPDF**

pyFPDF指适用于Python的FPDF，PyFPDF是从PHP移植而来的Python下生成PDF文档的库，主要用于将HTML网页转换为PDF文件。

项目主页：https://github.com/reingart/pyfpdf

安装方法：

* 在线安装：pip install fpdf
* 本地安装（pyfpdf安装包.zip）：python setup.py install

代码已经封装好，只需执行以下代码即可完成将HTML网页转换为PDF文件：

html2pdf.html2pdf(html\_text,path)

1. **实验要求**

将给定的LaTex文件（example.tex）转换为PDF文件，流程如下：

* 基于正则表达式将LaTex源文件进行解析
* 将解析结果构建为HTML字符串
* 通过pyfPDF生成pdf

正则表达式

Latex源文件

HTML文件

PDF

FPDF

需完成以下标记的解析如下标记：

* document
* title
* abstract
* section
* subsection
* itemize/item
* tabular
* emph
* textbf

示例程序和example.tex位于example/ 文件夹。

**建议用时：**2~4课时