

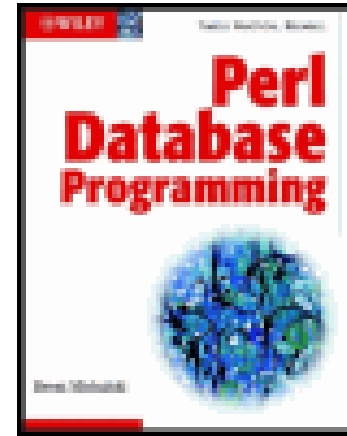
Perl 入门和提高 Lesson 9

周晓方

courses@xfzhou.homeftp.org

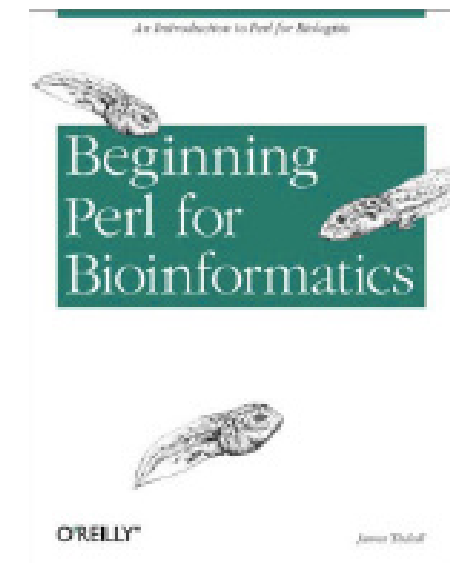
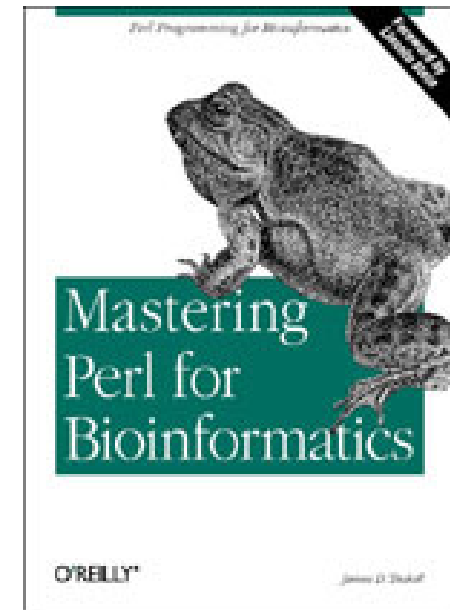
Perl和数据库

- 学习数据库和SQL
- 熟悉和安装MySQL
- 看书B.Michalski, "Perl Database Programming", *John Wiley & Sons Inc* 2003
- 看书A. Descartes, T. Bunce, "Programming the Perl DBI", *O'Reilly* 2000
- Perl的数据库往往和CGI以及Web Server一起工作
- PHP、Delphi在这方面用得也很广泛



Perl和生物信息学

- Bioinformatics
 - 需要大量反复地处理基因序列
- 参考书
 - James Tisdall, "Mastering Perl for Bioinformatics", O'Reilly 2003
 - James Tisdall, "Beginning Perl for Bioinformatics", O'Reilly 2001

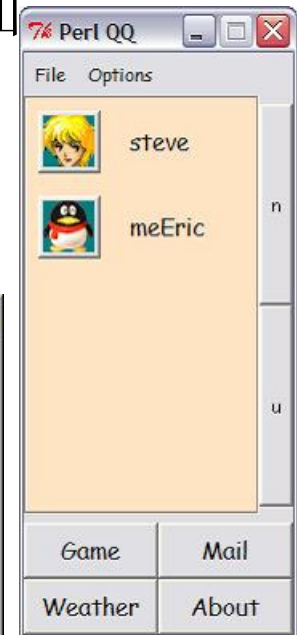


Perl和EDA

- TCL/TK用得较广泛，Perl在这方面还的书籍较少
- Perl作为字符处理很强的Glue Language，可以快速写些脚本，可以替代以前用Shell写的脚本
- 各公司内部EDA部门使用Perl
- 一些应用举例
 - bin文件和srec文件之间的转换，字节分割等
 - 自动生成Verilog格式的RTL状态机
 - 用perl生成冗长的testbench

Perl和Graph, GUI

- Perl产生gif文件, GD模块
- Perl和TK, 参考
 - <http://www.lns.cornell.edu/~pvhp/ptk/ptkTOC.html>
 - Perl的在线文档C:/Perl/html/site/lib/Tk.html
 - 《即时应用Perl模块》第6章,创建图形用户界面
 - 《Perl编程经典》第21章,图形化与Tk
 - Nancy Walsh, "Learning Perl/TK", O'Reilly 1999
 - Steve Lidie, Nancy Walsh, "Mastering Perl/TK", O'Reilly 2002



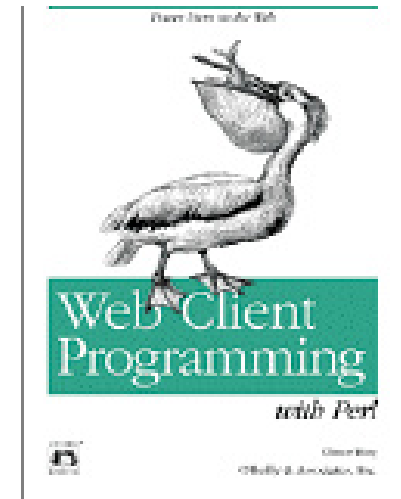
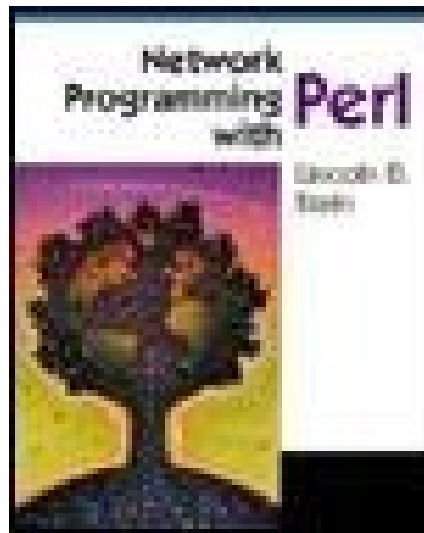
O'REILLY®



Mastering Perl/Tk



Perl和Win32



Perl和网络

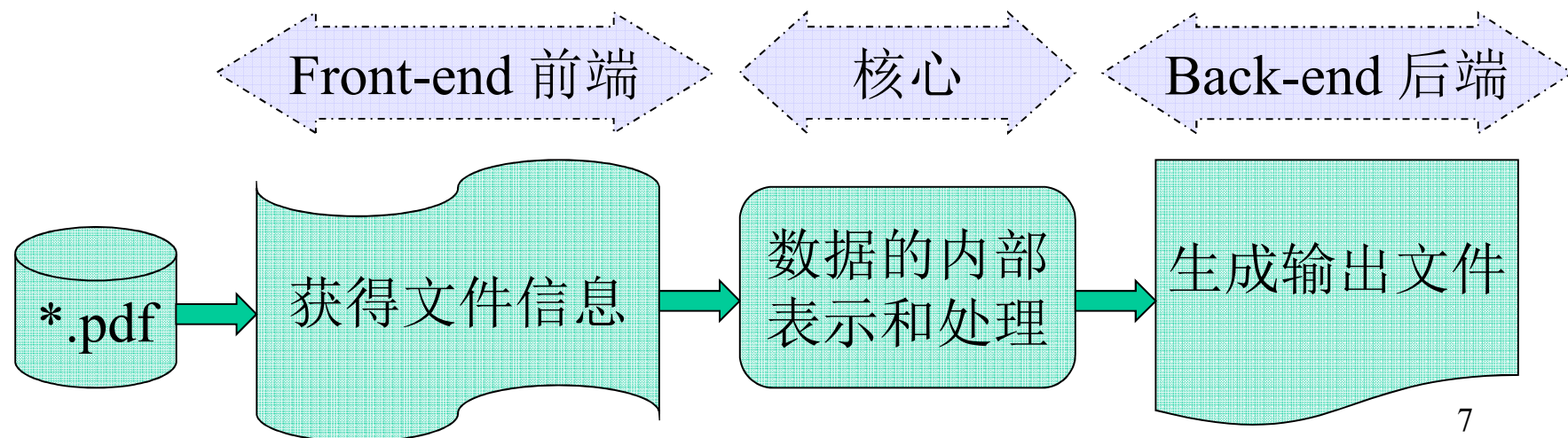
- perl编写网络服务器、客户端
- perl的CGI接口
- perl处理XML、HTML

使用CPAN解决问题-设计实例

- 问题的提出
 - Pdf文件列表
 - 读取*.pdf文件
 - 生成网页,罗列标题等信息
- 解决问题的思路何在

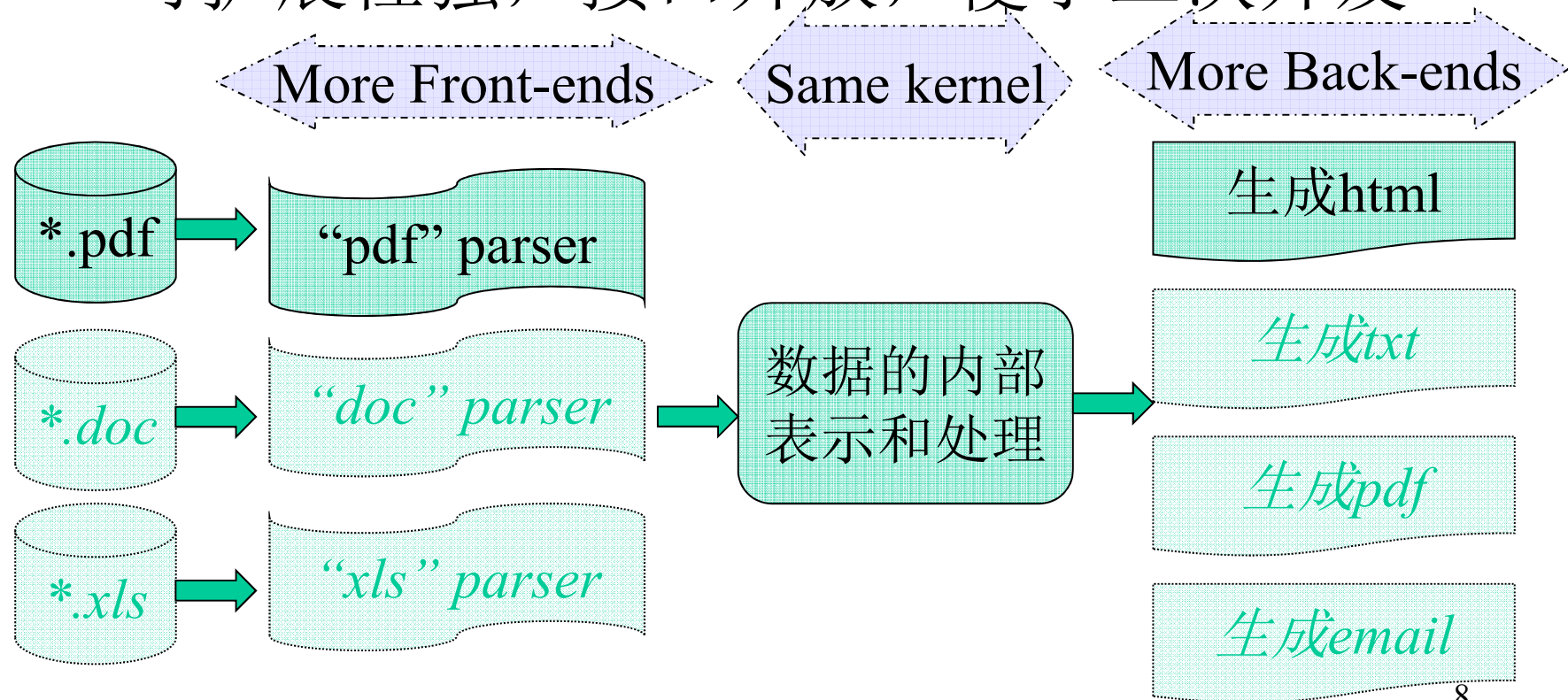
• PDF with a title: (filename / pages / PDF ModDate / titile information)			
00913285.pdf	6	2001-03-30	The multiple variable order problem for binary decision diagrams: theory and practical application - Design Automation Conference, 2001. Proceedings of the ASP-DAC 2001. Asia and South Pacific
• Bad PDF file: (filename / error message)			
00813186.pdf	Can't read cross-reference section, according to trailer		
00845087.pdf	Can't read cross-reference section, according to trailer		
00856407.pdf	Can't read cross-reference section, according to trailer		

pdftitle.pdf v1.7, Copyright (c) 2002 Xiaofang Zhou.
This is a copyrighted freeware still under developing.
Use it at your own risk. Bugs report to xfzhou@etang.com.



分解成前后端的好处

- 模块化，结构清晰
 - 模块内部要“内聚”，模块之间要“松散”
- 可扩展性强，接口开放，便于二次开发



解决方案

- 如何获得pdf文件的标题等信息？
 - 看看CPAN上有没有现成的模块可以用
 - 学习pdf格式，编写词法语法分析器(lex-yacc)
- 寻找、筛选相关的CPAN模块
 - PDF相关模块很多，我选用古老的PDF-111
- 数据有了，生成怎样的输出格式？
 - 生成文本的报告➔直接print
 - 生成网页的报告➔直接print或用CGI模块
 - 生成其他格式的文件➔OLE? CPAN? DIY?

编程和调试

- 阅读手册，安装模块
- 编写小脚本，测试模块的功能
- 直接在一个叫is_pdf的例子修改, 实现自己的脚本
- 发现的问题：
 - 有些pdf文件会整死模块和脚本→eval包装
 - 无法解析有些pdf→给出错误信息, 先不解决