

标题: 【锦书工厂】- 自有书管理,SNB阅读,TXT分章导入,PDF图文重排 [\[打印本页\]](#)

作者: zhuanyouzi 时间: 2011-3-4 14:51 标题: 【锦书工厂】- 自有书管理,SNB阅读,TXT分章导入,PDF图文重排

本帖最后由 zhuanyouzi 于 2011-3-8 22:33 编辑

既然进了前30，当然希望能有更多人使用它。

之前没有写过任何关于【锦书工厂】的帖子，所以有必要现在整理一下了。

写这个程序的初衷是，对于我个人来说，电脑中有很多 PDF 文件和 TXT 文件，我希望能把这些都转化成 SNB 格式到 Bambook 中阅读，在转换成 SNB 之前，我希望能预先阅读一下，也希望这个工具能对 PC 中和 Bambook 中的图书一目了然方便管理。取名【锦书工厂】是希望它像工厂一样源源不断生产出 SNB 格式的电子书。可能 RSS 什么的对很多媒体或者IT人士来说有很大的吸引力,但是我想更多的广大 Bambook 用户只是希望能有个好用的导入工具吧。

【主要功能】

- PC端个人书库管理。包括：PDF导入、TXT章节剪刀、添加和删除SNB。
- PC端个人书库与Bambook同步。
- PC端的电子书阅读器，不需要用Bambook就可以在PC上直接看书。

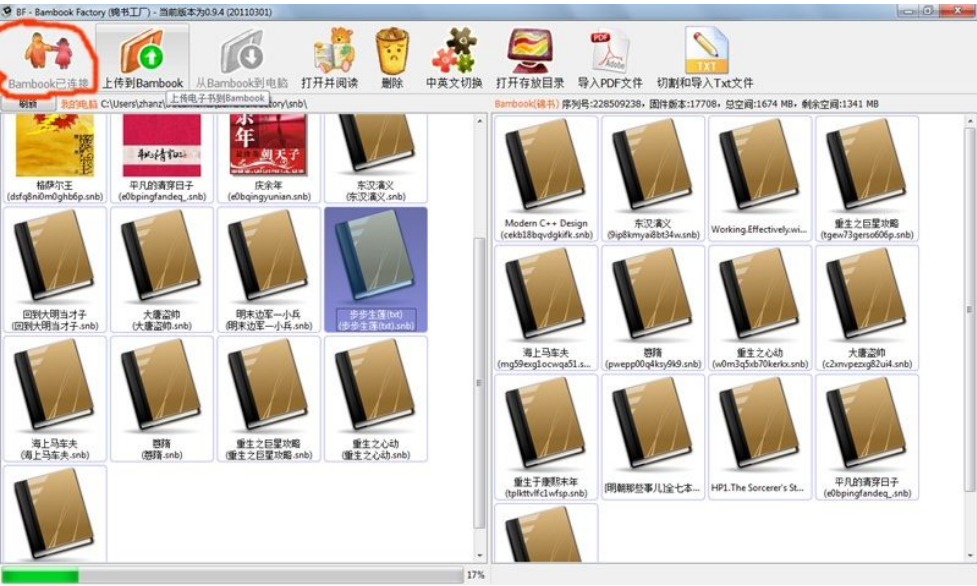
【特点】

- 稳定的程序运行
- 使用多线程自动检测锦书连接状态
- 卡通简易的操作界面，打开可立即使用，无需查看任何帮助
- 类似真实书本的阅读界面带来良好的用户体验，可在同步之前先试读一下
- 支持多语言切换，在全英文操作系统中正确显示中文
- 正确安装运行时库，没有测出任何兼容性问题，如UAC，64位兼容性问题，乱码问题，保证程序稳定运行
- 集成PDF终结者功能，图片和文字分别提取，重排
- 集成TXT剪刀终结者功能，智能分章，并可编辑分章，分章结果不满意时，可以放宽条件再次智能分章

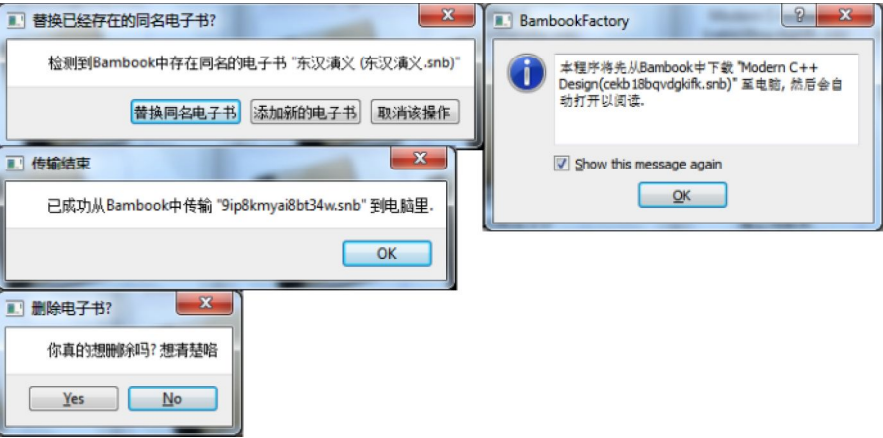
【功能 1】PC端个人书库同步和管理(添加 / 删除)

程序可在锦书未连接的情况打开，同时程序启动一个背景线程，检测锦书连接状态，并自动更新连接状态。

主程序界面，这是正在上传文件至锦书时的截屏：



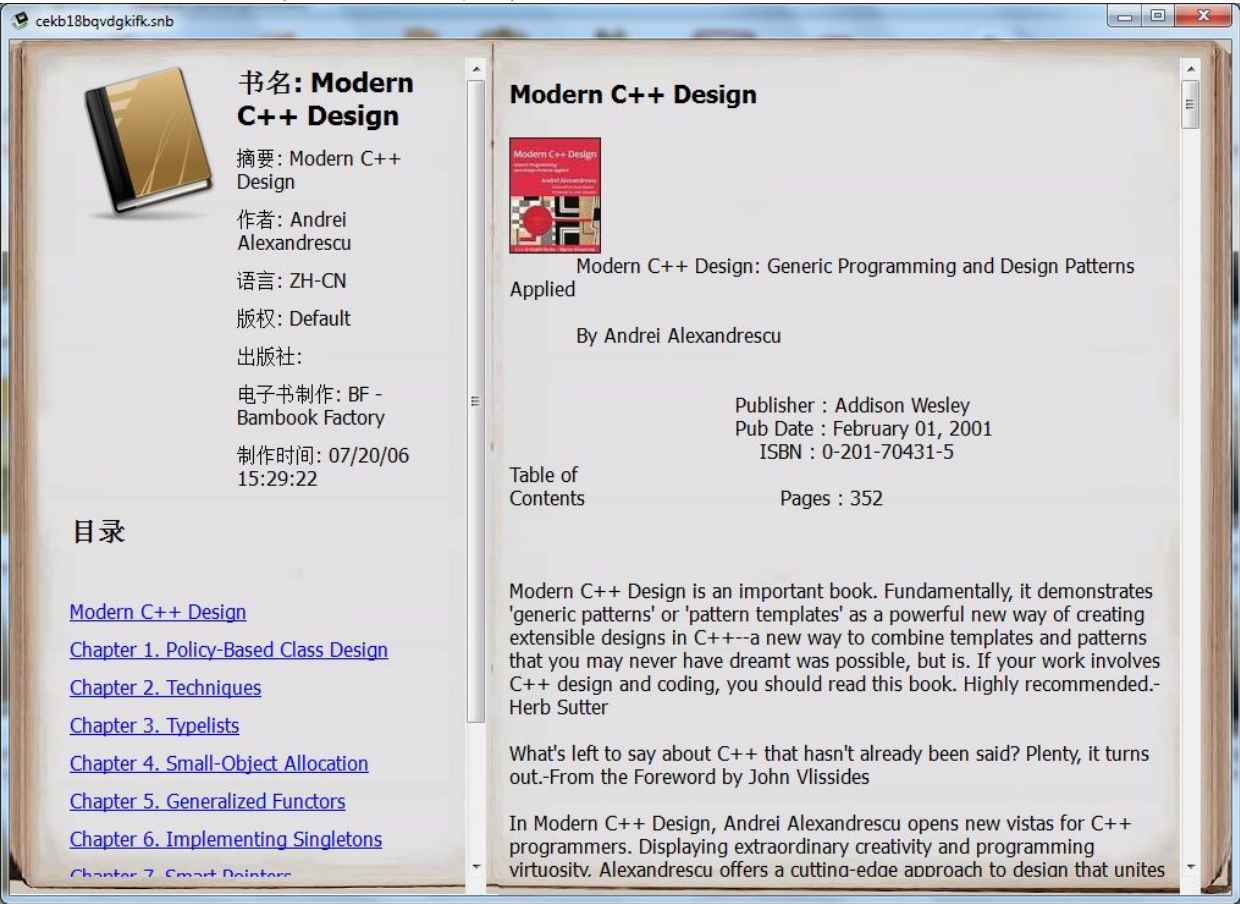
上传 / 下载 / 删除等情况下的一些对话框：



【功能 2】SNB图书阅读器

这是本人比较满意的一个界面，同时对它的性能有很大信心，即使打开单章几M大小的文本，速度也很快。

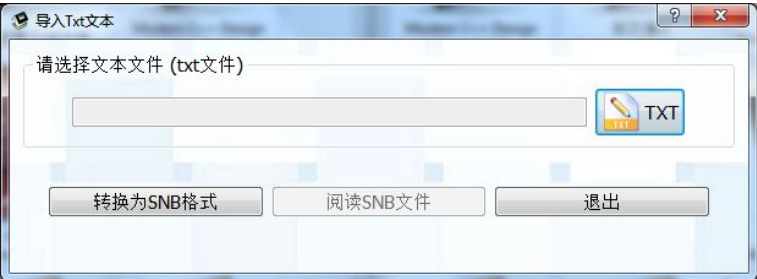
下图为从PDF中导入的SNB文件（图片和文字分别提取，重排）。



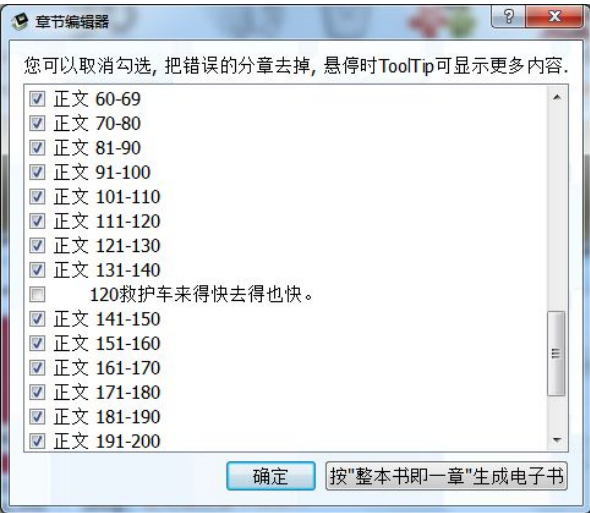
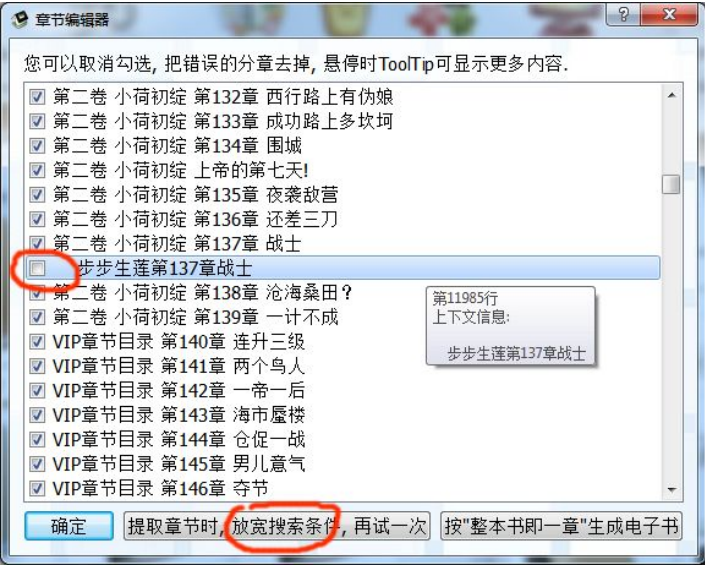
【功能 3】TXT转SNB

- 1. 智能分章
- 2. 可编辑智能分章的结果
- 3. 如果智能分章结果不正确, 可以选择放宽条件重新分章
- 4. 可以整本书只生成一章

大概【锦书工厂】宣传力度不够(程序员的通病?), 我很惊奇地发现有的软件光有这个小功能就有很多人使用并且得到很高的评价。因此我觉得这个子功能在同类软件中也是相当出色的。

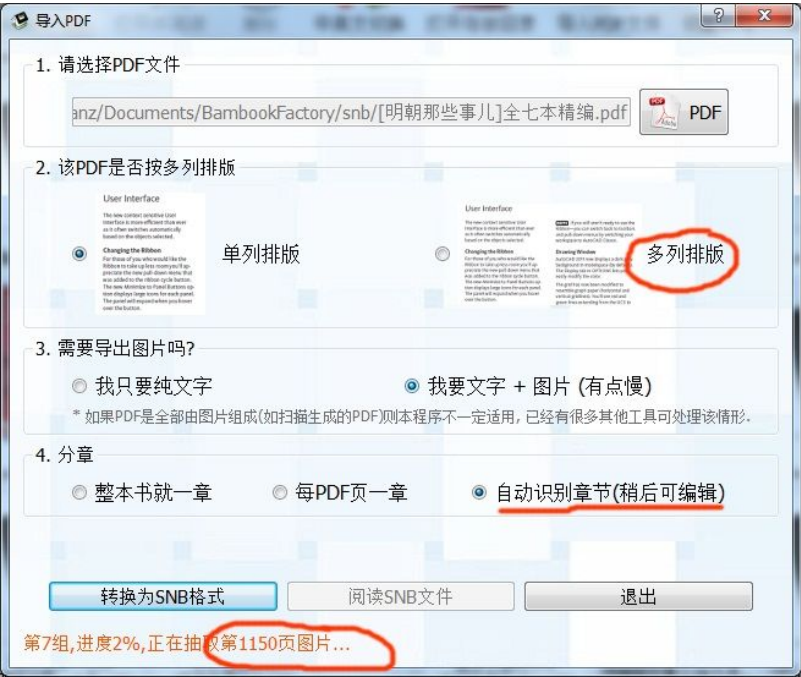


提取章节时,可以有多重提取算法,对当前提取结果不满意时,可以采用更为宽松的提取算法,  
第一重,提取 "第xxx(章|节|集|卷|部|回)" 的标题文本  
第二重,提取 " xxx(章|节|集|卷|部|回)" 的标题文本  
第三重,提取 " 1 xxx" 的标题文本



**【功能 4】PDF转SNB**


1. 图片和文字分别提取, 重排
2. 智能处理文本换行
3. 智能处理目录
4. 智能分章
5. 可编辑智能分章的结果
6. 可处理多列的PDF



上图是一个正在提取文本和图片的过程中。在实现提取图片时，为了提高速度，创建了10个线程同时提取不同页面段的图片。

以下是一些例子，导入前后效果对比图：

1. PDF截图 - Modern C++ Design.pdf



**Modern C++ Design: Generic Programming and Design Patterns Applied**  
By [Andrei Alexandrescu](#)

Publisher : Addison Wesley  
Pub Date : February 01, 2001  
ISBN : 0-201-70431-5  
Pages : 352

[Table of Contents](#)

*Modern C++ Design* is an important book. Fundamentally, it demonstrates 'generic patterns' or 'pattern templates' as a powerful new way of creating extensible designs in C++--a new way to combine templates and patterns that you may never have dreamt was possible, but is. If your work involves C++ design and coding, you should read this book. Highly recommended.-Herb Sutter

What's left to say about C++ that hasn't already been said? Plenty, it turns out.-From the Foreword by John Vlissides

In *Modern C++ Design*, Andrei Alexandrescu opens new vistas for C++ programmers. Displaying extraordinary creativity and programming virtuosity, Alexandrescu offers a cutting-edge approach to design that unites design patterns, generic programming, and C++, enabling programmers to achieve expressive, flexible, and highly reusable code.

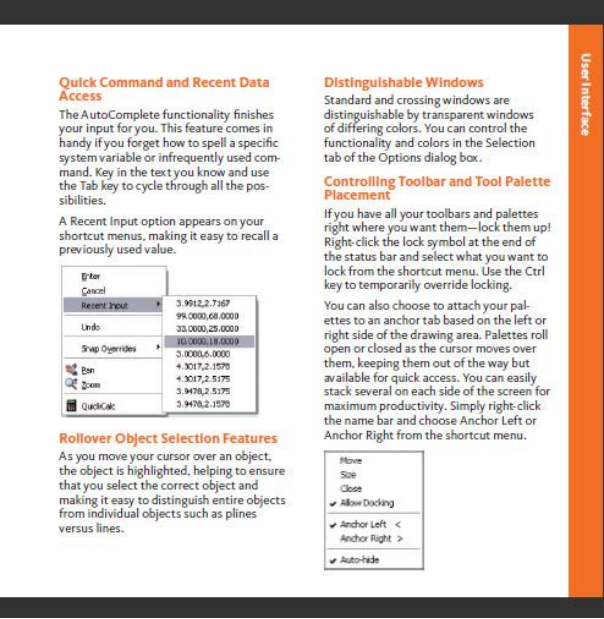
This book introduces the concept of generic components-reusable design templates that produce boilerplate code for compiler consumption-all within C++. Generic components enable an easier and more seamless transition from design to application code, generate

SNB效果截图。只提取文本，注意自动处理换行的结果，自动删除前面空格的效果，以及保留格式的效果:

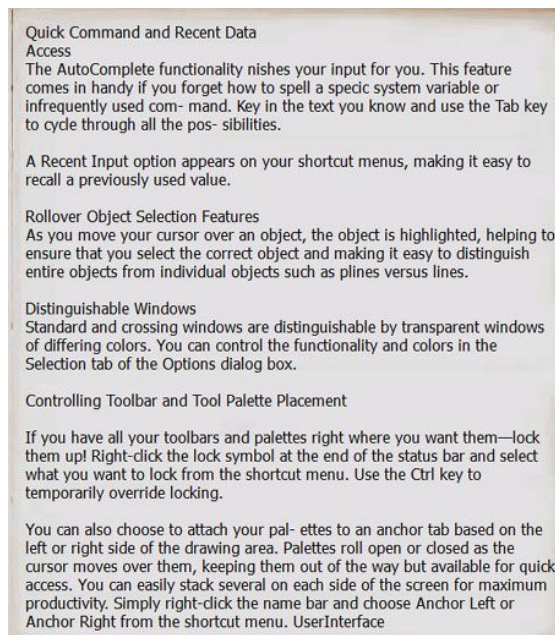




2. 多列 PDF 文件截图



SNB 效果截图, 注意文本重新排版的效果:



3. PDF里中英文混合的情况.

## Dive Into Python

2004 年 5 月 20 日

译文版 (5.4) : 2005 年 12 月—2006 年 4 月 (update-060425)

审校 (5.4b) : 2007 年 6 月—9 月

Copyright © 2000, 2001, 2002, 2003, 2004 Mark Pilgrim  
(<mailto:mark@diveintopython.org>)

Copyright © 2001, 2002, 2003, 2004, 2005, 2006, 2007 CPyUG (邮件列表)  
(<mailto:python-cn@googlegroups.com>)

本书存放在 <http://diveintopython.org/> (英文原版) 和  
<http://www.woodpecker.org.cn/diveintopython/> (中文版)。如果你是从别的地方  
看到它的, 可能看到的不是最新版本。

Permission is granted to copy, distribute, and/or modify this document under  
the terms of the GNU Free Documentation License, Version 1.1 or any later  
version published by the Free Software Foundation; with no Invariant Sections,  
no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included  
in [Appendix G, GNU Free Documentation License](#).

允许在 GNU 自由文档协议 (1.1 版, 或自由软件基金会出版的任何更新版本)  
的许可下复制、发行且/或修改本文档; 本文档没有不变部分, 没有前封面文  
本, 没有封底文本。该协议的一份中文版参考译文包含在 [Appendix H, GNU  
自由文档协议](#) 中。

SNB效果图, 即使中英文混合也能正确处理换行, 正确重排。



4. 另外,PDF中目录索引通常太长,对BAMBOOK来说会产生很多换行影响阅读。所以对这种情况也进行了处理:

Table of Content.....i  
Copyright.....vi  
Foreword.....vii  
Foreword.....ix  
Preface.....x  
Audience.....xi  
Loki.....xi  
Organization.....xii  
Acknowledgments.....xiii  
Part I: Techniques.....1  
Chapter 1. Policy-Based Class Design.....2  
1.1 The Multiplicity of Software Design.....2  
1.2 The Failure of the Do-It-All Interface.....3  
1.3 Multiple Inheritance to the Rescue?.....4  
1.4 The Benefit of Templates.....5  
1.5 Policies and Policy Classes.....6  
1.6 Enriched Policies.....9  
1.7 Destructors of Policy Classes.....10  
1.8 Optional Functionality Through Incomplete Instantiation.....11  
1.9 Combining Policy Classes.....12  
1.10 Customizing Structure with Policy Classes.....13  
1.11 Compatible and Incompatible Policies.....14  
1.12 Decomposing a Class into Policies.....16  
1.13 Summary.....17  
Chapter 2. Techniques.....19  
2.1 Compile-Time Assertions.....19

【锦书工厂】大的功能主要就这4个（个人书库同步和管理，SNB图书阅读器，TXT自动分章和导入，PDF智能导入）

欢迎大家使用，并且给出意见和建议。

欢迎BUG REPORT。



图片附件: main.jpg (2011-3-4 14:51, 140.88 KB) / 下载次数 11  
<http://bambookbbs.sdo.com/attachment.php?aid=5690&k=ed563f85449076965fbb05bf78a5ed69&t=1317884987>