

# 项目演示内容

- "经文"一览
- ●"人物"一览
- ●"兵器"一览
- "妖怪" 一览
- ●"问题"一览



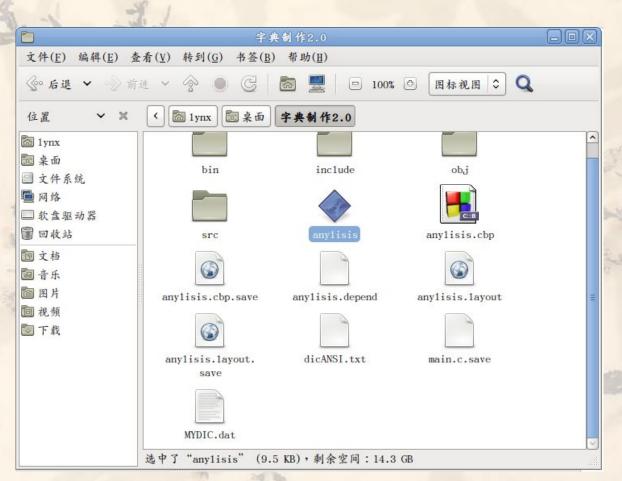
# "经文"

"经文" 1——Penguin

|                         | Penguin注音软件 |          |                                      |            |
|-------------------------|-------------|----------|--------------------------------------|------------|
| E文件 ¥视图                 |             |          |                                      | <u>H</u> 养 |
| 青在这里输入:                 | 这里显示        | 结果:      |                                      |            |
| 欠迎来到Penguin软件发布现场!      | huan1<br>欢  | ying2 la | Penguin ruan3<br>Penguin 软<br>!<br>! | jian4<br>件 |
|                         |             |          |                                      |            |
|                         |             |          |                                      |            |
|                         |             |          |                                      |            |
|                         |             |          |                                      |            |
|                         |             |          |                                      |            |
| Penguin程序由Penguin项目组所有! |             |          | 打开 注:                                | 音 保存       |

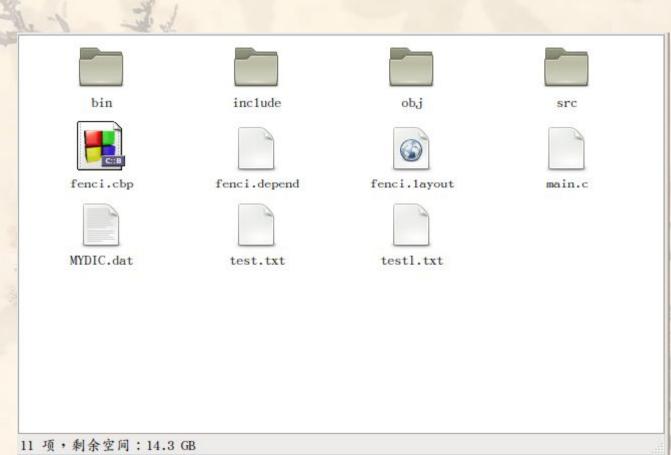
# "经文"

#### "经文"2——字典生成程序



# "经文"

"经文"3——字典缺失文字检查



### "人物"

- "人物"——项目组成员
- ●唐僧——项目目标,每个人都有自己的目标
- ●悟空——能力很强,但是时间精力有限
- ●八戒——可以为项目组发现不少问题,为项目组带来快乐
- ●沙僧——默默无闻但是非常给力的完成了项 目要求
- ●龙马——直接贡献少但是完成了不可缺少的 任务

#### "兵器"

## "兵器"——设计方法与思路一: code::blocks

```
Watches

✓ Local variables

text = 0x0

ans = 0x0

saveSuccess = 0

✓ Function Arguments

argc = 1

argv = 0xbffff944

✓ root

wordSequance = 0xb7bb9008

sequanceLength = 25382

maxLength = 11
```

```
src/pyGraphic.c | src/mydic.c | src/main.c |
 Projects Symbols
                                               #include "../include/mydic.h"
 ▽ Workspace
                                        29
                                               #include "../include/participle.h"
                                        30
  ▽ 40.3
                                               indexes *root = NULL;
                                        31
    ▽ Sources
                                        32
                                               int main(int argc,char *argv[])
      ▼ 🗁 src
                                        33
                                        34
          accounting.c
                                                   char *text = NULL;
                                        35
          assertion.c
                                                   char *ans = NULL;
                                        36
                                                   BOOL saveSuccess = FALSE;
                                        37
                                        38
          mydic.c
                                                   setlocale(LC ALL, "");
                                        39
          participle.c
                                        40
          pyFile.c
                                                   root = openMyDic();
                                        41
          pyGraphic.c
                                        42
                                                   if (argc != 3)//图形模式
                                        43
    ▶ B Headers
                                        44
                                        45
                                                       tuxing(argc, argv);
                                        46
Scripting console
                                        47
                                                   else//命令行模式
Welcome to the script console!
                                        48
                                        49
                                                       text = readfile(argv[1], strlen(argv[1]));
                                        50
                                        51
                                                       if (text == NULL)
                                        52
                                        53
                                                           printf("File didn not read in!\n");
                                        54
                                                           abort();
                                        55
```

#### "兵器"

"兵器"——设计方法与思路: 词库分析 通过将读入的字典Trie树保存为一个二进制文 件后, 在我的T5500 (1.6GHz双核CPU) 上读 入字典时间是0.141s。这样如果字库有任何改 变,只要更改DicANSI.txt文件(UTF-8格式) 运行一次anylisis后即可生成新的MyDic.dat文 件。将它复制到字典文件检查缺失汉字的目录 下,对一篇很长的文字检查后,save.txt保存着 没有添加的单字。这样我们又需要更新字典, 来完善程序功能。

#### "兵器"

"兵器"——设计方法与思路:交叉编译

在win平台下可以安装MinGW和GTK,这样代码可以在win平台和Ubuntu下不加修改的编译,非常方便测试和调试。

# "妖怪"

"妖怪"——BUG 保存使用wchar\_t型,Ubuntu内无法打开, windows可以打开查看,结果正确。 这是因为winxp使用GB2312,Ubuntu使用 UTF-8。可以改变系统的编码,但是setlocale 可以获得当前系统的编码,不用修改即可适应 当前的系统编码。

### "妖怪"

"妖怪"——BUG 若输入的不是文本而是路径,则提示段错误 这个bug修改很简单,就是判断打开的路径后, 对于无效路径不反应即可



# "妖怪"

"妖怪"——BUG

动态分配、变长数组在运行中导致出错,出错位置为libc.o的第6行。

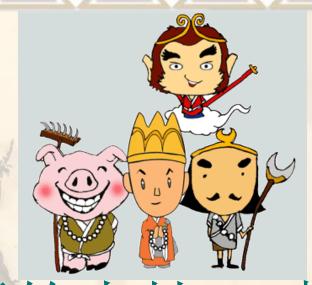
这个错误让人很费解,出现的情况是当free的 时候或者定义变长数组的时候。目前的解决办 法是使用定长数组。

### "问题"

"问题"——遗留的难题

#### 第2音无法选择

选择多音字词的第二音需要对该字词的属性或者前后文字判断,这个过程需要更多的资料,语料库的训练,还有程序自学习的能力。关键是如何突破纯粹的统计来达到分词的突破,还是一个难点。因此没有完成。



感谢老师的支持,感谢李艺组 无私的帮助,尤其感谢全部组 员的努力,我们最终才能达到 我们的目标!谢谢!