闲话汇编与接口

Tips on Assembly Language and Interface

张华平

Email: kevinzhang@bit.edu.cn

Website: http://www.nlpir.org/

@ICTCLAS张华平博士

大数据搜索与挖掘实验室

课程钉钉群

- 不得发跟课程无关的信息
- www.nlpir.org Resource资源\Teaching教学 获取汇编与接口 课件
- 平时成绩20/30分
  - 平时作业
  - 上机：必选2道；
  - 4人组队：编制小游戏
- 期末闭卷考试80/70

从Apple密码门看汇编...

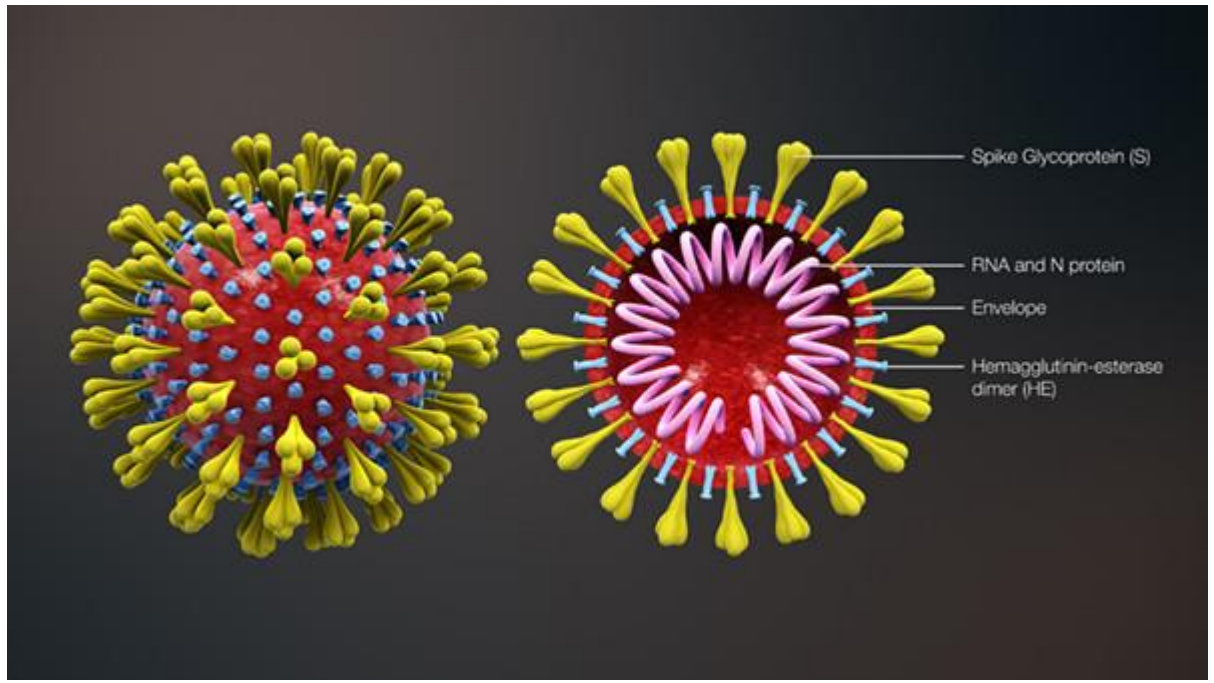
多语种背后的0与1

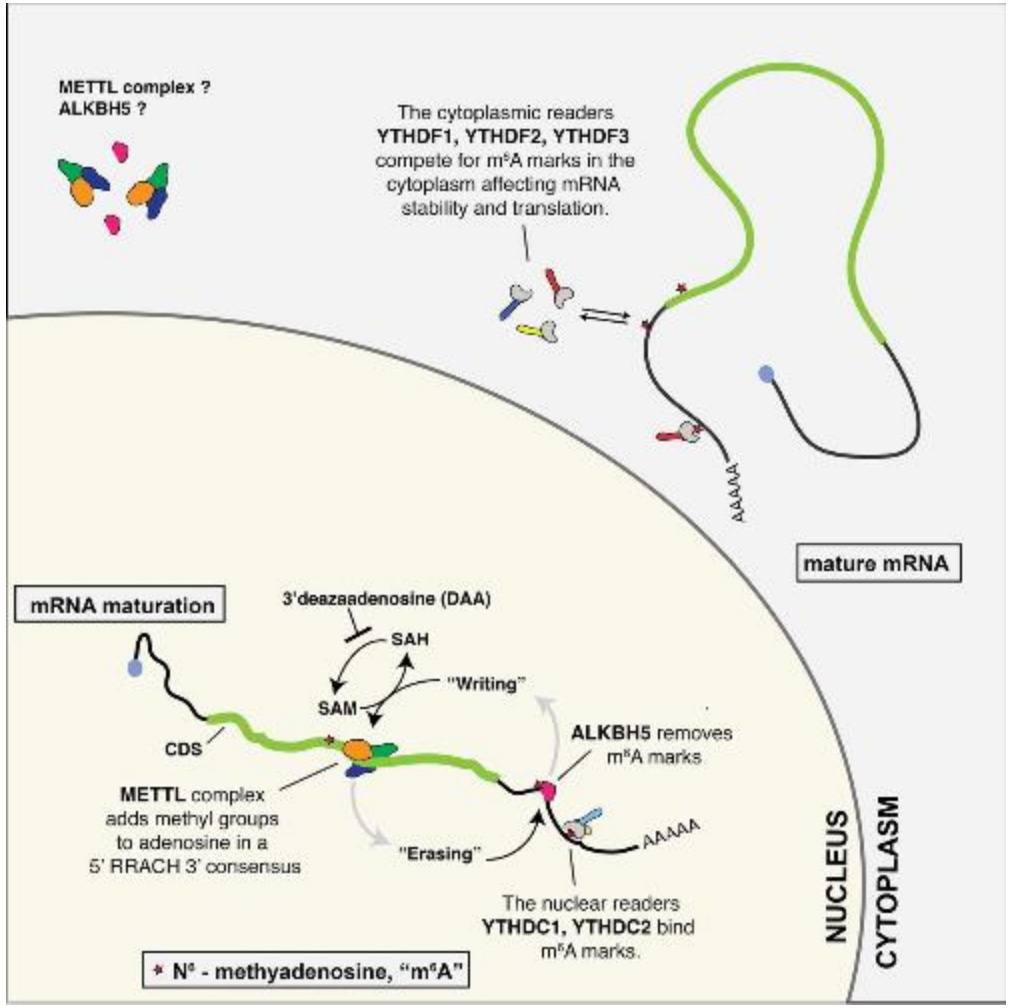
/MLIP
多语种智能信息处理
Intelligent MultiLingual Information Processing

多语种背后的0与1

/ML/P
多语种智能信息处理
Intelligent MultiLingual Information Processing

从新冠病毒防疫看汇编...

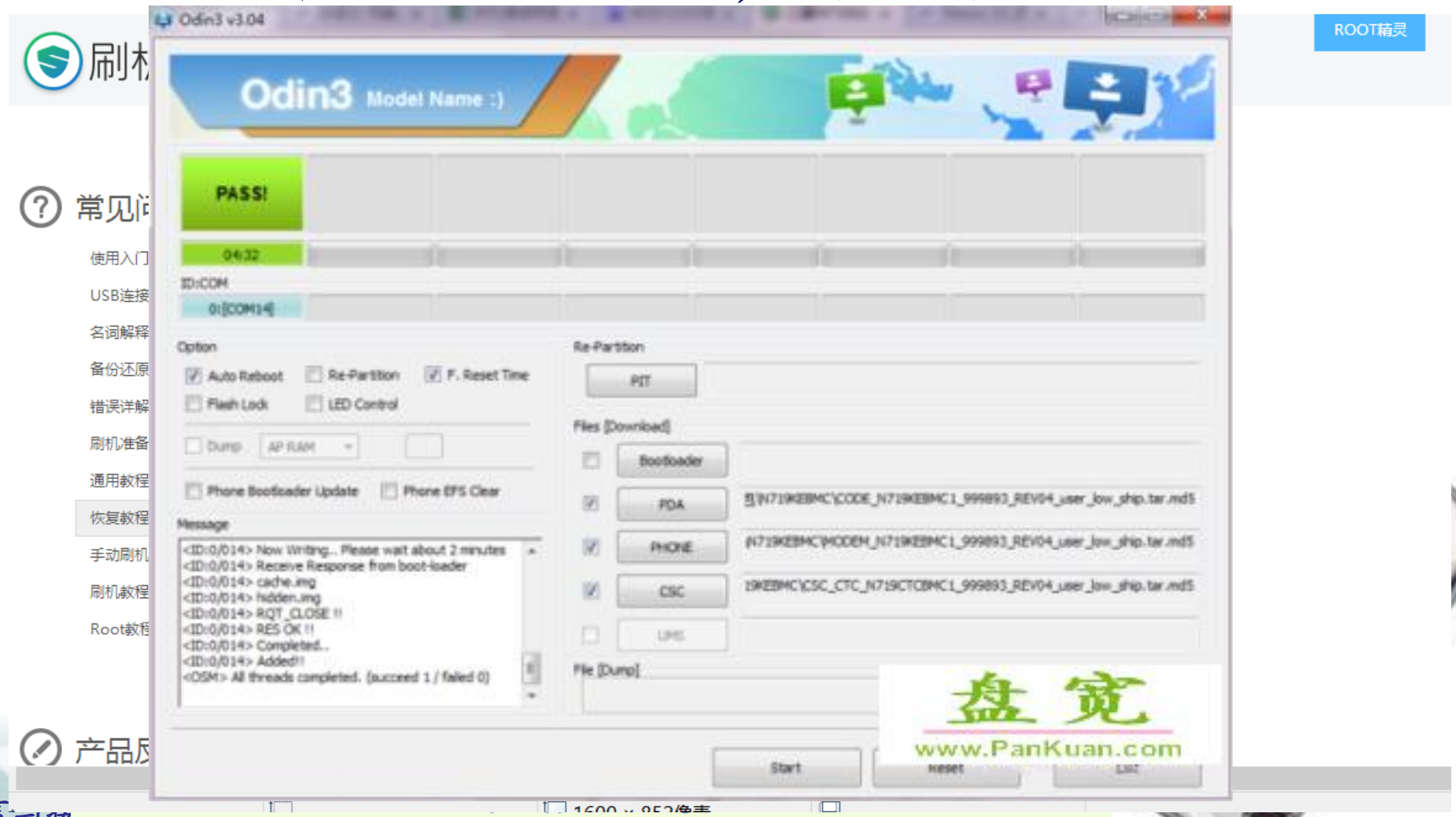




Kennedy, E. M., Courtney, D. G., Tsai, K., & Cullen, B. R. (2017, May 1). Viral Epitranscriptomics. Retrieved from https://jvi.asm.org/content/91/9/e02263-16

合约机的坑里木有汇编老师...

- 大家掉过合约机的坑，有木有？



闲话教育

兴趣第一

- 感兴趣找方法，不感兴趣找借口；
- 教育第一原则是培养对科学或者具体学科的兴趣，扼杀青年的兴趣，罪莫大焉；
- 再好的学问，以面目可憎的形象出现，年轻人也不可能接受。佛家无色无相，却幻化万象，以渡众生。

闲话教育

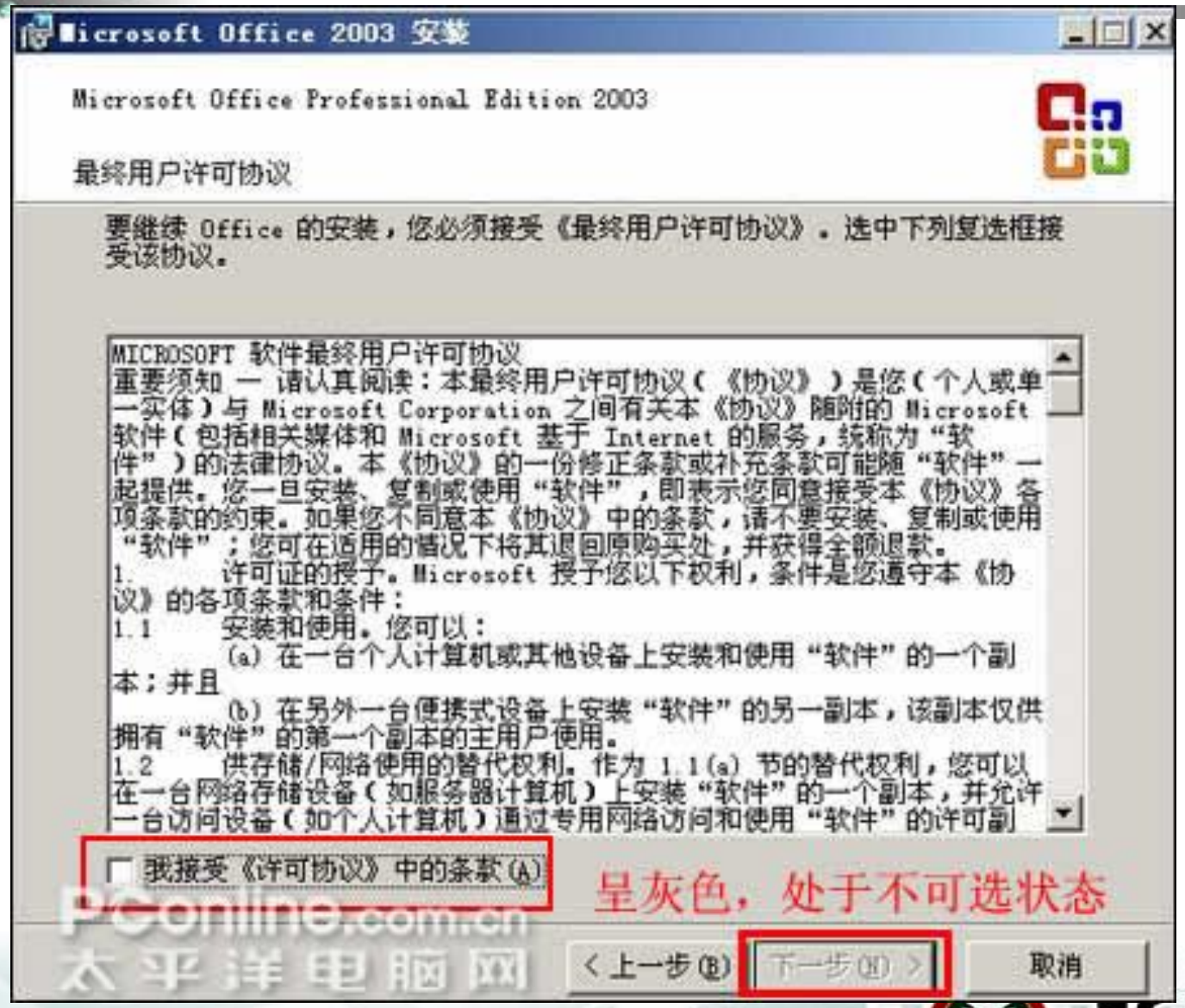
知行合一

- 明 王守仁《传习录》卷 教育家：陶行知
- 王守仁，号阳明先生，中国明代最著名的思想家、哲学家、文学家和军事家。陆王心学之集大成者，非但精通儒家、佛家、道家，而且能够统军征战，是中国历史上罕见的全能大儒。封“先儒”，奉祀孔庙东庑第58位。
- 计算机科学尤其强调知行合一。

闲话汇编接口

- 汇编不会编；其实汇编都会编；
- 计算机专业与非专业之三大差别：数据结构、汇编、编译；
  - 数据结构：解决问题的算法-原理；
  - 编译：软件如何编译，如何优化；
  - 汇编：软件更懂硬件；
- 1946年计算机发明以来，过去了60多年，高级语言已经可见即可得，为什么还要学汇编？

软件许可协议



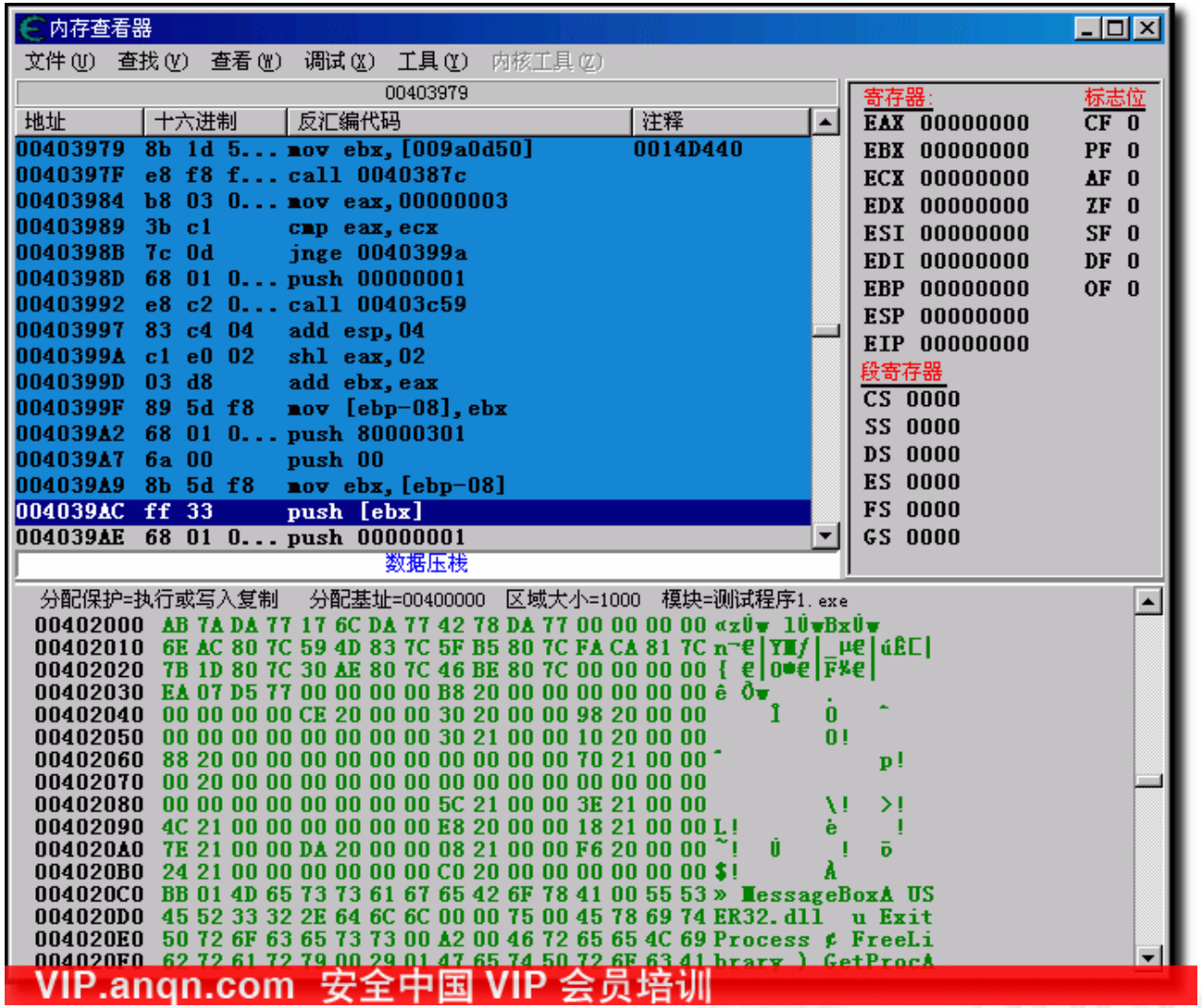
为什么都有下面这一条？

4. 对反向工程、反编译和反汇编的限制。您不得对"软件"进行反向工程、反编译或反汇编，除非适用法律明示允许，并仅在适用法律明示允许的范围内从事上述活动。

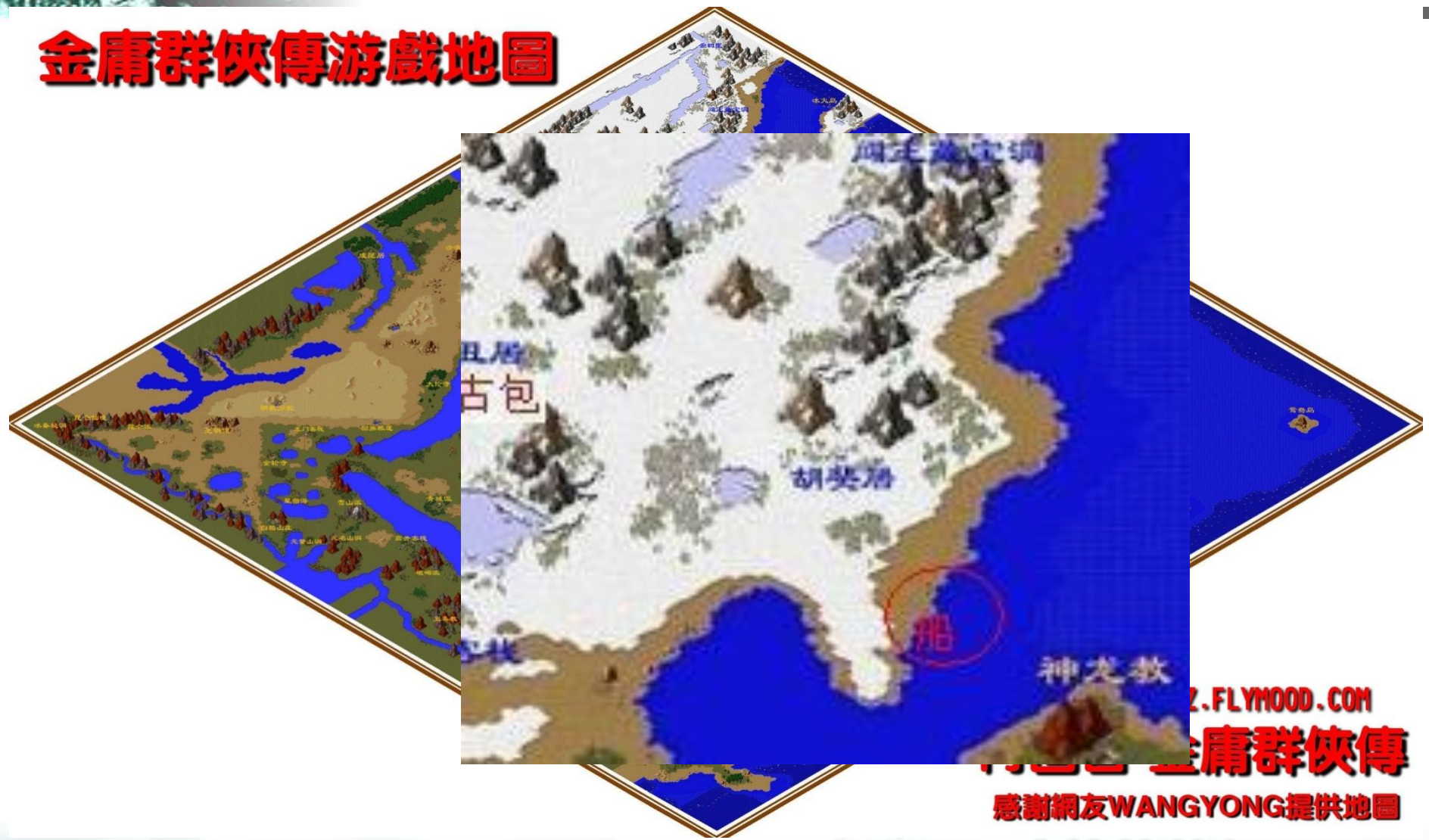
debug

```
08048414 <main>:
 8048414:       55                      push   %ebp
 8048415:       89 e5                   mov    %esp,%ebp
 8048417:       83 e4 f0                and    $0xfffffff0,%esp
 804841a:       83 ec 20                sub    $0x20,%esp
 804841d:       c7 44 24 1c 03 00 00    movl   $0x3,0x1c(%esp)
 8048424:       00
 8048425:       8b 44 24 1c             mov    0x1c(%esp),%eax
 8048429:       01 c0                   add    %eax,%eax
 804842b:       83 44 24 1c 01          addl   $0x1,0x1c(%esp)
 8048430:       03 44 24 1c             add    0x1c(%esp),%eax
 8048434:       89 44 24 18             mov    %eax,0x18(%esp)
 8048438:       83 44 24 1c 01          addl   $0x1,0x1c(%esp)
 804843d:       83 44 24 1c 01          addl   $0x1,0x1c(%esp)
 8048442:       b8 30 85 04 08          mov    $0x8048530,%eax
 8048447:       8b 54 24 18             mov    0x18(%esp),%edx
 804844b:       89 54 24 08             mov    %edx,0x8(%esp)
 804844f:       8b 54 24 1c             mov    0x1c(%esp),%edx
 8048453:       89 54 24 04             mov    %edx,0x4(%esp)
 8048457:       89 04 24                mov    %eax,(%esp)
 804845a:       e8 e1 fe ff ff          call   8048340 <printf@plt>
 804845f:       c7 04 24 00 00 00 00    movl   $0x0,(%esp)
 8048466:       e8 e5 fe ff ff          call   8048350 <exit@plt>
 804846b:       90                      nop
```

win-debug工具


内存查看器
文件(U) 查找(V) 查看(W) 调试(X) 工具(Y) 内核工具(Z)
00403979
地址 十六进制 反汇编代码 注释
00403979 8b 1d 5... mov ebx,[009a0d50] 0014D440
0040397F e8 f8 f... call 0040387c
00403984 b8 03 0... mov eax,00000003
00403989 3b c1 cmp eax,ecx
0040398B 7c 0d jnge 0040399a
0040398D 68 01 0... push 00000001
00403992 e8 c2 0... call 00403c59
00403997 83 c4 04 add esp,04
0040399A c1 e0 02 shl eax,02
0040399D 03 d8 add ebx,eax
0040399F 89 5d f8 mov [ebp-08],ebx
004039A2 68 01 0... push 80000301
004039A7 6a 00 push 00
004039A9 8b 5d f8 mov ebx,[ebp-08]
004039AC ff 33 push [ebx]
004039AE 68 01 0... push 00000001
数据压栈
寄存器: 标志位
EAX 00000000 CF 0
EBX 00000000 PF 0
ECX 00000000 AF 0
EDX 00000000 ZF 0
ESI 00000000 SF 0
EDI 00000000 DF 0
EBP 00000000 OF 0
ESP 00000000
EIP 00000000
段寄存器
CS 0000
SS 0000
DS 0000
ES 0000
FS 0000
GS 0000
分配保护=执行或写入复制 分配基址=00400000 区域大小=1000 模块=测试程序1.exe
00402000 AB 7A DA 77 17 6C DA 77 42 78 DA 77 00 00 00 00
00402010 6E AC 80 7C 59 4D 83 7C 5F B5 80 7C FA CA 81 7C
00402020 7B 1D 80 7C 30 AE 80 7C 46 BE 80 7C 00 00 00 00
00402030 EA 07 D5 77 00 00 00 00 B8 20 00 00 00 00 00 00
00402040 00 00 00 00 CE 20 00 00 30 20 00 00 98 20 00 00
00402050 00 00 00 00 00 00 00 00 30 21 00 00 10 20 00 00
00402060 88 20 00 00 00 00 00 00 00 00 00 00 70 21 00 00
00402070 00 20 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00402080 00 00 00 00 00 00 00 00 5C 21 00 00 3E 21 00 00
00402090 4C 21 00 00 00 00 00 00 E8 20 00 00 18 21 00 00
004020A0 7E 21 00 00 DA 20 00 00 08 21 00 00 F6 20 00 00
004020B0 24 21 00 00 00 00 00 00 C0 20 00 00 00 00 00 00
004020C0 BB 01 4D 65 73 73 61 67 65 42 6F 78 41 00 55 53
004020D0 45 52 33 32 2E 64 6C 6C 00 00 75 00 45 78 69 74
004020E0 50 72 6F 63 65 73 73 00 A2 00 46 72 65 65 4C 69
VIP.anqn.com 安全中国 VIP 会员培训

金庸群侠传的故事



密码破译的故事

```
00007f70h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
00007f80h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
00007f90h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
00007fa0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
00007fb0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
00007fc0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
00007fd0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
00007fe0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
00007ff0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
00008000h: 00 00 00 00 00 00 00 00 00 00 00 00 C1 1C 40 00 ; ............?@.
00008010h: B4 42 40 00 00 00 00 00 00 00 00 00 66 1D 40 00 ; 碆@.........f.@.
00008020h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
00008030h: 59 6F 75 20 61 72 65 20 69 6E 76 61 6C 69 64 21 ; You are invalid!
00008040h: 0A 00 00 00 48 65 6C 6C 6F 20 57 6F 72 6C 64 21 ; ....Hello World!
00008050h: 0A 00 00 00 59 6F 75 20 61 72 65 20 76 61 6C 69 ; ....You are vali
00008060h: 64 21 0A 00 42 49 54 00 25 73 00 00 50 6C 65 61 ; d!..BIT.%s..Plea
00008070h: 73 65 20 69 6E 70 75 74 20 70 61 73 73 77 6F 72 ; se input passwor
00008080h: 64 3A 0A 00 E9 26 40 00 01 00 00 00 20 09 2D 0D ; d:..?@..... .-.
00008090h: 5D 00 00 00 5D 00 00 00 60 AE 40 00 00 00 00 00 ; ]...]...`胤.....
000080a0h: 60 AE 40 00 01 01 00 00 00 00 00 00 00 00 00 00 ; `胤............
000080b0h: 00 10 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
000080c0h: 00 00 00 00 02 00 00 00 01 00 00 00 00 00 00 00 ; ................
000080d0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
000080e0h: 00 00 00 00 02 00 00 00 02 00 00 00 00 00 00 00 ; ................
000080f0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
00008100h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
00008110h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; ................
```

病毒如何检测识别？

- 基于机器学习的计算机恶意程序检测模型构建与实现-胡巍巍 学号：20082892
- 可执行程序转换为汇编代码，进行训练，特征识别，分类识别。

种类	频率	相对频率	提高
后门	96.63%	96.79%	0.16%
构造器	94.49%	94.92%	0.43%
木马	94.65%	95.88%	1.24%
病毒	97.23%	97.24%	0.01%
蠕虫	95.00%	95.73%	0.73%
其他	94.45%	95.12%	0.67%
平均	95.41%	95.95%	0.54%

NLPIR多语种认知智能团队

定位： 世界一流·多语种·认知智能·创新中心

20年多语种语义技术积累

- ✓ **中文信息处理最高奖**钱伟长一等奖（国内唯一分词方向）
- ✓ **算法：** 20+自主可控全链路NLP模块
  20M 缓存边缘计算
  数据： 10GB语料库，20亿语义知识库
  知识： 10+行业先验知识积累

组装人事物时空智能分析

- ✓ **多模态融合：** NLP+OCR+语音+图文比对的语义增强分析平台
- ✓ **启动：** <100份 小样本冷启动
  分析： KGB知识图谱关联分析
  生成： 报告智能生成

巡场订阅模式

- ✓ **国内：** 400+ 本地部署标杆客户
  军工、中央网信办、公安部、国研中心、人行、建行、中电科、航天科工、国家电网、华为等
- ✓ **全球：** 40万用户验证，新闻集团、韩国RSN、意大利Expert.AI、新加坡南洋理工、日立等

北京理工大學
BEIJING INSTITUTE OF TECHNOLOGY

张华平 博士

- ICTCLAS汉语分词创立者

  创建并运营NLPIR大数据语义增强分析平台

- 北京理工大学副教授，大数据搜索与挖掘实验室主任

✓ 中文信息处理领域最高奖：钱伟长中文信息处理一等

（全国唯一“分词”方向）

✓ 新疆自治区科技进步二等奖

✓ 第一届ACL-SIGHAN国际汉语分词大赛，国家973汉语评测第一，国际TREC Novelty国际第一

✓ 中国人工智能学会多语种信息智能处理专委会秘书长，中国中文信息学会社会媒体处理专委会副秘书长。

✓ 中央网信办、中宣部、公安部、国办电子政务总体组、军委某部规划评审专家

✓ “十三五” ZF第一个AI项目课题组长，“十四五”QB规划论证专家；jun科研多个领域战略情报

钱伟长一等奖证书

新疆科技进步二等奖

张华平教授受CCTV采访解读苹果FBI揭秘大战。

北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

基于NLPIR的多语种认知智能分析平台

场景层

- 防务：JS情报、科技情报、智能CM、文书核查、公文生成、辅助定密、战争量化、配置安全
- 网信宣传：内容过滤、市长信箱、扫黄打非、网信监测
- 公检法纪监：警情分析、WW预测、重复报警、手机侦查
- …：…、…

产品层

- 九眼智能情报平台
- 九芯智能文书平台
- NLPIR多语种语义智能分析平台（NLPIR语义分析引擎+JZSearch搜索+九眼过滤+KGB知识图谱工具等）

技术层

- 数据导入：精准采集、OCR、语音识别、DB导入、文档解析
- 多语分析：分词标注、依存文法、情感分析、语言模型、文本校对
- 信息抽取：新词发现、摘要提取、实体抽取、关键词提取、过滤与脱敏
- 文本挖掘：文档去重、相似比对、文本分类、文本聚类、话题分析
- 知识关联：全文检索、地图可视化、词向量扩展、知识生成、知识推理

基础层

- 算源：语料库、知识库、外部库
- 算法：专家启发、机器学习、终身学习
- 算力：国产芯片、国产OS、可控云

NLPIR inside

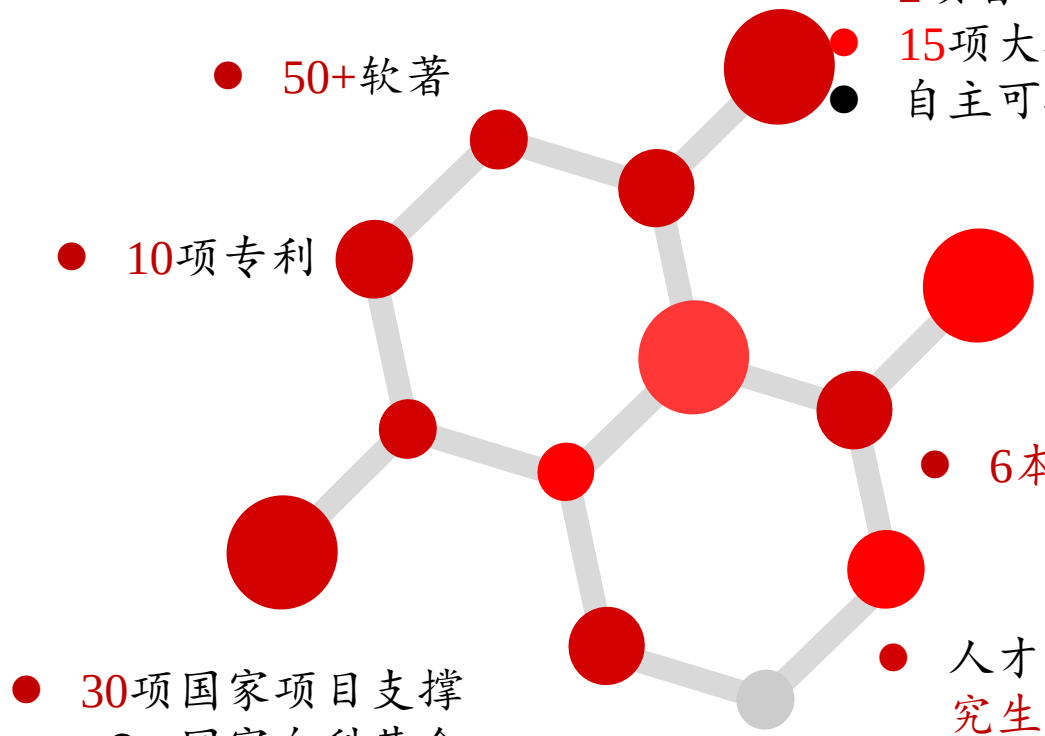
语义增强分析平台

- ✓ 具备70-80%应用功能
- ✓ 一周实现目标测试
- ✓ <100份样本冷启动
- ✓ 降低20-30%开发成本
- ✓ 提升系统智能化水平

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

主要成果

- 1个NLPIR大数据语义智能分析平台
- 2项省部级科技奖
- 15项大数据语义分析算法组件
- 自主可控，全面支持国产CPU/操作系统



- 50+软著
- 10项专利
- 100+学术论文，SCI/EI 30+
- 6本专著
- 人才培养：70+博士硕士研究生，网信十佳讲师1人
- 30项国家项目支撑
  - 国家自科基金
  - 国家重点研发
  - 国家242
  - ZF预研/GF创新特区
- 1000+论文引用
- 服务在线用户11.9亿人
- 近3年42万+技术开发用户
- **直接经济效益12.4亿元**

Thanks

Email: kevinzhang@bit.edu.cn
Weibo：@ICTCLAS张华平博士

Welcome to visit lab. homepage
http://www.nlpir.org