

部分开源工具和语料资源

1、NLTK 官方提供的语料库资源列表

http://nltk.googlecode.com/svn/trunk/nltk_data/index.xml

2、OpenNLP 上的开源自然语言处理工具列表

<http://opennlp.sourceforge.net/projects.html>

3、斯坦福大学自然语言处理组维护的“统计自然语言处理及基于语料库的计算语言学资源列表”

<https://nlp.stanford.edu/links/statnlp.html>

4、LDC 上免费的中文信息处理资源

<https://www ldc.upenn.edu/collaborations/past-projects>

课件

- 1、哈工大刘挺老师的“统计自然语言处理”课件；
- 2、哈工大刘秉权老师的“自然语言处理”课件；
- 3、中科院计算所刘群老师的“计算语言学讲义”课件；
- 4、中科院自动化所宗成庆老师的“自然语言理解”课件；
- 5、北大常宝老师的“计算语言学”课件；
- 6、北大詹卫东老师的“中文信息处理基础”的课件及相关代码；

以上 下载地址：

链接：<https://pan.baidu.com/s/1geIFdYF> 密码：lmt3

7、MIT 大牛 Michael Collins 的“Machine Learning Approaches for Natural Language Processing(面向自然语言处理的机器学习方法)”课件；

8、Michael Collins 的“Machine Learning (机器学习)”课件；

9、SMT 牛人 Philipp Koehn “Advanced Natural Language Processing (高级自然语言处理)”课件；

10、Philipp Koehn “Empirical Methods in Natural Language Processing”课件；

11、Philipp Koehn “Machine Translation (机器翻译)”课件。

书籍

- 1、《自然语言处理综论》英文版第二版
- 2、《统计自然语言处理基础》英文版
- 3、《用 Python 进行自然语言处理》，NLTK 配套书
- 4、《Learning Python 第三版》，Python 入门经典书籍，详细而不厌其烦

以上 下载地址：

链接：<https://pan.baidu.com/s/1hs8d0dY> 密码：e954

- 5、《自然语言处理中的模式识别》
- 6、《EM 算法及其扩展》
- 7、《统计学习基础》
- 8、《自然语言理解》英文版（似乎只有前 9 章）
- 9、《Fundamentals of Speech Recognition》，质量不太好，不过第 6 章关于 HMM 的部分比较详细，作者之一便是 Lawrence Rabiner;
- 10、概率统计经典入门书：
《概率论及其应用》（英文版，威廉*费勒著） 第一卷 第二卷 DjVuLibre 阅读器（阅读前两卷书需要）
- 11、一本利用 Perl 和 Prolog 进行自然语言处理的介绍书籍：《An Introduction to Language Processing with Perl and Prolog》
- 12、国外机器学习书籍之：
 - 1) “Programming Collective Intelligence “，中文译名《集体智慧编程》，机器学习&数据挖掘领域”近年出的入门好书，培养兴趣是最重要的一环，一上来看大部头很容易被吓走的”
 - 2) “Machine Learning “，机器学习领域无可争议的经典书籍，下载完毕将后缀改为 pdf 即可。
豆瓣评论 by 王宁）：老书，牛人。现在看来内容并不算深，很多章节有点到为止的感觉，但是很适合新手（当然，不能”新”到连算法和概率都不知道）入门。比如决策树部分就很精彩，并且这几年没有特别大的进展，所以并不过时。另外，这本书算是对 97 年前数十年机器学习工作的大综述，参考文献列表极有价值。国内有翻译和影印版，不知道绝版否。
 - 3) “Introduction to Machine Learning”
- 13、国外数据挖掘书籍之：
 - 1) “Data Mining. Concepts. and. Techniques. 2nd “，数据挖掘经典书籍。华裔科学家写的书，相当深入浅出。
 - 2) Data Mining: Practical Machine Learning Tools and Techniques
 - 3) Beautiful Data: The Stories Behind Elegant Data Solutions (Toby Segaran, Jeff Hammerbacher)
- 14、国外模式识别书籍之：
 - 1) “Pattern Recognition”
 - 2) “Pattern Recognition Technologies and Applications”
 - 3) “An Introduction to Pattern Recognition”
 - 4) “Introduction to Statistical Pattern Recognition”
 - 5) “Statistical Pattern Recognition 2nd Edition”
 - 6) “Supervised and Unsupervised Pattern Recognition”
 - 7) “Support Vector Machines for Pattern Classification”
- 15、国外人工智能书籍之：
 - 1) Artificial Intelligence: A Modern Approach (2nd Edition) 人工智能领域无争议的经典。

2) “Paradigms of Artificial Intelligence Programming: Case Studies in Common LISP”

16、其他相关书籍:

1) Programming the Semantic Web, Toby Segaran, Colin Evans, Jamie Taylor

2) Learning.Python 第四版, 英文