# 安利幾個優質nlp開源項目

公子龍 2022-03-22 13:10

### 1、OpenNRE

OpenNRE 是基於Tensorflow 開發的，一個用於神經網絡關係提取的工具包，由清華大學劉知遠老師及其團隊貢獻的開源項目。在該項目中，關係提取會分為嵌入、編碼器、選擇器和分類器四步。

Github 地址：

> https://github.com/thunlp/OpenNRE

### 2、中文序列標註Flat Lattice

原文《FLAT：Chinese NER Using Flat-Lattice Transformer》，解決的是中文命名實體識別的任務。文章提出的方法在多個數據集上達到了SOTA結果，目前是中文NER的一個主流的方法。

Github 地址：

> https://github.com/LeeSureman/Flat-Lattice-Transformer

### 3、文本分類

Github 地址：

> https://github.com/dennybritz/cnn-text-classification-tf

### 4、序列建模

Github 地址：

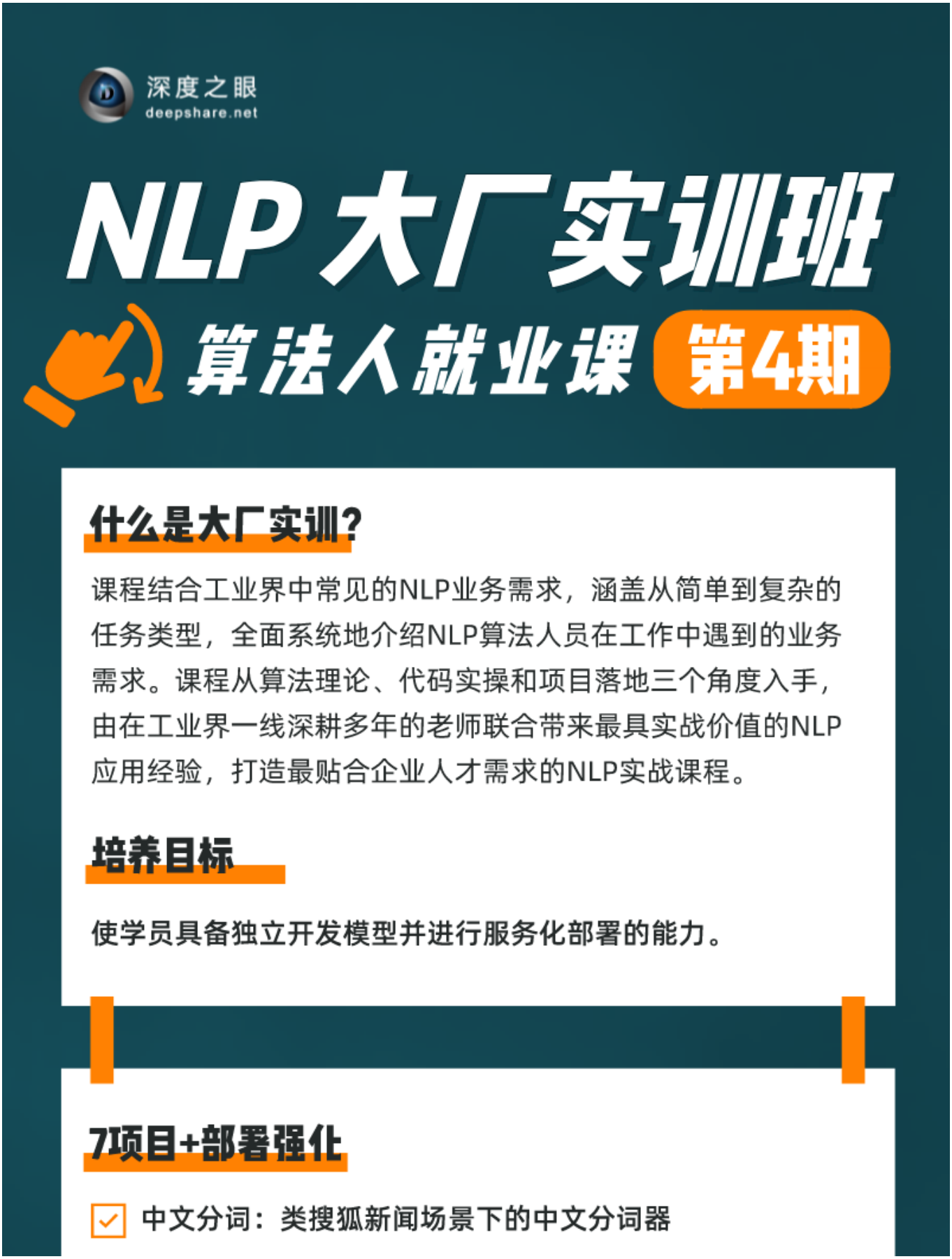
> https://github.com/google/seq2seq

### 5、中文分詞

Github 地址：

> https://github.com/koth/kcws

如果你覺得這些開源項目量級不夠或者自己學起來吃力，我推薦你參加深度之眼的NLP大廠實訓班，一線算法工程師帶你學習常見NLP項目。



- 关键词提取：类新浪门户场景下的关键词提取
- 实体识别：类新浪微博场景下的实体识别
- 文本分类：头条新闻标题分类场景下的BERT分类器训练、优化及蒸馏
- 文本摘要：从0到1实现一个摘要系统
- 对话系统：工业级对话系统项目
- 知识图谱项目实战：构建与应用
- 部署强化：工程化部署

**Step1.实战理论体系**

项目使用背景、算法历史沿革、模型原理讲解、多种算法比较

**Step2.工业项目实践**

真实典型数据、企业要求规范、项目解决方案，代码实操演示

**4步走直通大厂就业**

**Step3.企业级部署**

REST接口、多种测试、Docker微服务、自动化CI/CD、k8s的GPU集群

**Step4.面试强化**

掌握大厂的面试技巧，系统地梳理面试中常遇到的问题体系

## 中文分词

### 1、分词器初步开发-从机械切分到序列标注模型切分

第1节:快速构建基于语言模型的机械分词器，迈出企业级分词第一步

第2节:模型来了！基于CRF和BiLSTM-CRF的序列标注分词，大幅提升精度

### 2、分词器终极开发-算法融合与快速反馈badcase

第1节:来看看业界著名的分词工具，Jieba和HanLP源码剖析
第2节:融合机械分词和模型切分，加上badcase快速反馈机制，打造属于你的工业级分词器

## 关键词提取

### 1、开发无监督的关键词提取器

第1节:快速构建一个基于tfidf和textrank的关键词提取系统，迈出第一步
第2节:融入主题模型和新词发现技术，让你的系统效果提升一大截
第3节:有监督的关键词提取简介，序列标注与多标签分类

## 实体识别

### 1、传统的HMM、CRF的实体识别器

第1节:基于HMM和CRF的序列标注
第2节:构建使用HMM、CRF的实体识别器，支持词库管理，快速构建一个传统的成熟NER系统

### 2、升级，深度学习版的实体识别器

第1节:深度学习来了！使用IDCNN、LSTM来进行NER，提升系统精度
第2节:使用Bert及其变体，更上一层楼

### 3、一些实践上的干货，用上它们，让你的NER系统又快又好

第1节:HMM、CRF、Bert调参经验
第2节:数据增强、标签分布不均衡、loss选择、正则化、ONNX推理加速

## 文本分类

### 1、监督分类模型开发及优化

第1节:集成学习和DL分类
第2节:长文本和模型蒸馏

## 文本摘要

### 1、文本摘要项目

第1节:文本摘要任务综述、指标实现和讲解
第2节:抽取式文本摘要、抽取式文本摘要模型实现和讲解
第3节:生成式文本摘要、模型实现和讲解
第4节:预训练摘要模型生成、基于预训练文本摘要模型实现和讲解
第5节:文本要模-进阶思考、相关模型的实现和讲解
第6节:模拟面试

## 对话系统

### 1、对话系统构建基础

第1节:系统性地介绍对话系统的结构和技术细节的发展历程
第2节:传统语义匹配算法

### 2、对话系统构建进阶

第1节:对比学习与最近邻召回
第2节:介绍任务型对话rasa框架

### 3、对话系统构建高级

第1节:系统性地介绍闲聊型对话系统的结构和技术细节的发展历程
第2节:模拟面试

## 知识图谱

### 1、知识图谱项目实战：构建与应用

第1节:知识图谱简介（图数据库neo4j简介）
第2节:实体识别模型：序列标注 vs. Biaffine
第3节:实体识别模型代码
第4节:关系抽取与指代消歧
第5节:关系抽取与指代消歧代码
第6节:实体(术语)标准化：系统搭建，模型实现
第7节:基于知识库的推荐系统
第8节:KGAT代码复现

## 工程部署

### 1、基于Docker微服务的NLP服务部署

第1节:服务的HTTP封装；单元测试、接口测试的编写；GPU的使用
第2节:Docker容器化服务；自动化CICD；负载均衡与服务扩缩容；K8S启动与GPU的调度

## 主讲老师

**王老师 算法研究员**

在金融、媒体、电商等领域有众多NLP项目落地项目。对微服务、GPU集群等领域也有丰富的实践经验。

**胡老师 高级算法工程师**

主持参与过对话系统、热搜榜等项目落地
在MLNLP、CIKM等顶级学术会议以一作身份发表过多篇论文

**张老师 大厂算法工程师**

5年多NLP相关工作经验，在医疗领域信息抽取、电商搜索推荐以及通用对话系统构建等领域有实际项目落地经验。

**Michael老师 算法研究员**

主持参与过多项业界落地项目
在MLNLP、CIKM等顶级学术会议发表过多篇顶会论文

本項目課程開源了3小時的項目嚐鮮課程供大家學習，我專門為本公眾號粉絲申請到了50個試學名額。

> 1、NLP工業場景
> 2、基於語言模型的機械切分
> 3、預處理-splitter、normalize
> 4、詞庫加載、詞圖構建

**原價~~399~~元，本號粉絲，僅0.1元！**