

# 自然语言处理应用开发 职业技能等级标准

标准代码：510086

（2021年2.0版）

北京京东世纪信息技术有限公司 制定

2021年12月 发布

# 目 录

前言.....	1
1 范围.....	2
2 规范性引用文件.....	2
3 术语和定义.....	2
4 适用院校专业.....	3
5 面向职业岗位 ( 群 ) .....	3
6 职业技能要求.....	4
参考文献.....	10

# 前 言

本标准按照GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本标准起草单位：北京京东世纪信息技术有限公司、天津城建大学、天津中德应用技术大学、北京信息职业技术学院、重庆电子工程职业学院、苏州市职业大学、重庆青年职业技术学院、东营职业学院、安徽职业技术学院、青岛职业技术学院、上海电子信息职业技术学院、威海海洋职业学院、深圳职业技术学院、黑龙江能源职业学院、南京信息职业技术学院、上海市信息技术学校、中国商业股份制企业经济联合会。

本标准主要起草人：任成元、雷玫、钱曦、戴红光、庞青、刘毅、孙锋、史宝会、武春岭、李金祥、唐中剑、周连兵、李京文、常中华、朱咏梅、童红兵、王隆杰、王路权、聂明、王鹤、孙文升。

声明：本标准的知识产权归属于北京京东世纪信息技术有限公司，未经北京京东世纪信息技术有限公司同意，不得印刷、销售。

## 1 范围

本标准规定了自然语言处理应用开发职业技能等级对应的工作领域、工作任务及职业技能要求。

本标准适用于自然语言处理应用开发职业技能培训、考核与评价，相关用人单位的人员聘用、培训与考核可参照使用。

## 2 规范性引用文件

下列文件对于本标准的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本标准。凡是不注日期的引用文件，其最新版本适用于本标准。

GB/T 5271.34-2006 人工智能神经网络

GB/T 5271.31-2006 人工智能机器学习

## 3 术语和定义

GB/T 5271.34-2006、GB/T 5271.31-2006界定的以及下列术语和定义适用于本标准。

### 3.1 人工智能 Artificial Intelligence

人工智能是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。

### 3.2 机器学习 Machine Learning

机器学习是人工智能的一门分支学科，主要研究算法并开发程序或系统，用于根据输入数据构建预测模型。这种系统能利用训练得到的模型对同一分布中提取的新数据进行实用的预测。

### 3.3 深度学习 Deep Learning

深度学习是机器学习的分支，是一种以人工神经网络为架构，对数据进行表征学习的算法。

### 3.4分词 word segmentation

分词就是将用自然语言书写的文章、句、段等，经电子计算机处理后，以词为单位给以输出，为后续加工处理提供先决条件。

### 3.5词性标注 Part-Of-Speech tagging

词性标注也被称为语法标注或词类消疑，是语料库语言学中将语料库内单词的词性按其含义和上下文内容进行标记的文本数据处理技术。

### 3.6 自然语言处理 Natural Language Processing

自然语言处理研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。自然语言处理是一门融语言学、计算机科学、数学于一体的科学。

## 4 适用院校专业

### 4.1 参照原版专业目录

中等职业学校：计算机应用、物联网技术应用、电子与信息技术、电子技术应用、软件与信息服务等专业。

高等职业学校：电子信息工程技术、物联网应用技术、人工智能技术服务、计算机应用技术、软件技术、嵌入式技术与应用、大数据技术与应用、计算机信息管理、云计算技术与应用等专业。

高等职业教育本科学校：电子信息工程、大数据技术与应用、软件工程等专业。

应用型本科学校：人工智能、智能科学与技术、信息安全、计算机科学与技术、软件工程、物联网工程、电子信息工程、电子科学与技术等专业。

## 4.2 参照新版职业教育专业目录

中等职业学校：计算机应用、大数据技术应用、物联网技术应用、电子信息  
技术、电子技术应用、软件与信息服务等专业。

高等职业学校：电子信息工程技术、物联网应用技术、人工智能技术应用、  
计算机应用技术、软件技术、嵌入式技术与应用、大数据技术、云计算技术与应  
用等专业。

高等职业教育本科学校：电子信息工程技术、大数据工程技术、软件工程技  
术、云计算技术、人工智能工程技术、嵌入式技术等专业。

应用型本科学校：人工智能、智能科学与技术、信息安全、计算机科学与技  
术、软件工程、物联网工程、电子信息工程、电子科学与技术等专业。

## 5 面向职业岗位（群）

主要面向包含自然语言处理相关业务的互联网企业、传统企事业单位等的自  
然语言处理应用产品研发、系统集成、测试运维以及解决方案部门，从事数据标  
注、智能产品训练、自然语言处理应用开发、系统测试、算法调优等工作任务。  
面向的主要岗位包括自然语言处理工程师、数据标注师、人工智能算法测试员  
（NLP 方向）、自然语言处理研发工程师、自然语言处理算法工程师等。

## 6 职业技能要求

### 6.1 职业技能等级划分

自然语言处理应用开发职业技能等级分为三个等级：初级、中级、高级，三  
个级别依次递进，高级别涵盖低级别职业技能要求。

【自然语言处理应用开发】（初级）：主要面向各类开展自然语言处理，物  
联网，智能机器人，数字服务等业务的企事业单位，从事文本数据的采集、清洗、

分词、标注等预处理工作，能根据项目测试计划和测试用例完成测试并形成测试报告，能根据项目要求对模型进行应用部署。

【自然语言处理应用开发】（中级）：主要面向各类开展自然语言处理，物联网，智能机器人，数字服务等业务的企事业单位，从事文本数据的采集、清洗、分词、标注等预处理工作及数据预处理质量控制，能够使用云服务商提供的接口进行自然语言处理相关的云服务应用开发，能够对基于机器学习的模型进行训练及推理应用。

【自然语言处理应用开发】（高级）：主要面向各类开展自然语言处理，物联网，智能机器人，数字服务等业务的企事业单位，从事指导验收数据预处理、系统架构、解决方案规划、模型选型等工作，完成机器学习模型的建模、训练、调优、效果评估和工程应用开发的任务。

## 6.2 职业技能等级要求描述

表 1 自然语言处理应用开发职业技能等级要求（初级）

工作领域	工作任务	职业技能要求
1. 语料采集与整理	1.1 语料采集	1.1.1 具备将纸质资料进行扫描形成电子化语料库的能力。 1.1.2 具备在国内外标准开放数据集（比如搜狗语料、人民日报语料等）中收集语料形成语料库的能力。 1.1.3 具备使用适当的工具进行语料采集或从网络爬取文本等数据的能力
	1.2 语料整理	1.2.1 具备能够根据实际需求，对段落、空格、字符等相关问题进行整理的能力。
2. 语料预处理	2.1 语料清洗	2.1.1 完成语料清洗的常用工具及质量的要求。 2.1.2 具备能够根据实际需求，选择合适的工具进行语料清洗的能力。
	2.2 分词模型训练	2.2.1 完成分词相关的常用工具及质量的要求。 2.2.2 具备能够根据实际需求，选择合适的工具进行分词的能力

工作领域	工作任务	职业技能要求
	2.3 词性标注与命名实体识别	2.3.1 完成词性标注相关的常用工具及质量的要求。 2.3.2 具备能够根据实际需求，选择合适的标注方法和工具进行标注的能力。
3.应用项目测试	3.1 测试计划制定	3.1.1 完成项目功能及需求的要求。 3.1.2 具备能够根据项目计划与开发人员沟通，制定测试的内容、规则、环境等测试计划文档的要求。
	3.2 测试用例编写	3.2.1 具备能够根据需求设计编写测试用例的能力。 3.2.2 达到熟练掌握 Excel 等常用文档编辑工具的条件。 3.2.3 能够根据实际需求，选择合适的测试工具。
	3.3 执行测试	3.3.1 能够严格按测试用例执行测试。 3.3.2 能够对开发人员修复 bug 后提交的新版本进行版本控制。 3.3.3 具备处理缺陷处理流程，提交缺陷，跟踪缺陷至缺陷关闭的能力。
	3.4 测试报告编写	3.4.1 能够根据测试结果编写测试报告。
4.应用场景部署	4.1 应用场景认知	4.1.1 完成文本分类、情感分析、智能写作、智能分诊等应用场景的学习。
	4.2 应用部署	4.2.1 能够根据特定应用场景，完成数据预处理和应用部署的工作。 4.2.2 能够根据特定应用场景，完成测试数据准备，应用效果验证的工作。

表 2 自然语言处理应用开发职业技能等级要求（中级）

工作领域	工作任务	职业技能要求
1. 语料采集与整理	1.1 语料采集	1.1.1 能够根据需要，设定采集数据的格式并保证采集数据的质量。 1.1.2 能够根据需要，选择合适的数据采集工具或通过 python 编写爬取脚本。
	1.2 语料整理	1.2.1 能够根据实际需求，对段落、空格、字符等相关问题进行整理。 1.2.2 能够根据实际需求，制定语料整理质量要求。 1.2.3 能够根据质量要求，对整理结果进行质量审核。
2. 语料预处理	2.1 语料清洗	2.1.1 能按照模型训练要求，制定语料清洗的方式和质量要求。 2.1.2 能够根据实际需求，选择合适的工具进行语料清洗。 2.1.3 能够根据质量要求，对清洗结果进行质量审核。



工作领域	工作任务	职业技能要求
	2.2 分词模型训练	2.2.1 能按照模型训练要求，制定分词的方式和质量要求。 2.2.2 能够根据实际需求，选择合适的分词工具进行分词处理。 2.2.3 能够根据质量要求，对分词结果进行质量审核。
	2.3 词性标注与命名实体识别	2.3.1 能按照模型训练要求，制定词性标注的方式和质量要求。 2.3.2 能够根据实际需求，选择合适的标注方法和工具。 2.3.3 能够根据质量要求，对标注结果进行质量审核。
3.自然语言处理云服务应用开发	3.1 云服务认知学习	3.1.1 完成云厂商提供的自然语言处理相关云服务接口的学习。 3.1.2 根据实际项目需求，能够选择合适的云服务接口进行应用开发。 3.1.3 根据实际项目需求，通过自定义工具，能够完成文本分类、智能写作、情感分析等符合自身要求的云服务。
	3.2 云服务接口调用	3.2.1 能够根据云服务接口说明，正确完成云服务账号的配置。 3.2.2 能够根据云服务接口说明，完成正确的参数设置。 3.2.3 能够根据云服务接口说明，正确完成接口调用。
	3.3 云服务结果处理	3.3.1 能够根据云服务接口说明，正确解析返回的结果。 3.3.2 能够根据云服务接口说明，处理异常返回。 3.3.3 能够根据反馈结果，完成结果的可视化展示。
4.深度学习模型训练与应用	4.1 机器学习模型训练	4.1.1 完成机器学习模型训练方法的学习。 4.1.2 完成有监督学习模型方法的学习。 4.1.3 完成有监督学习模型的多种评估方法的学习。
	4.2 深度学习模型认知	4.2.1 完成 TensorFlow 开发环境的学习。 4.2.2 完成卷积神经网络的人脸识别技术的学习。 4.2.3 完成循环神经网络的门控算法学习。 4.2.4 完成长短时记忆网络的算法学习。
	4.3 深度学习模型训练	4.3.1 完成深度学习框架 TensorFlow、Mxnet 等开发环境的学习。 4.3.2 能够根据要求，完成模型训练的数据准备。 4.3.3 能够根据要求，迭代数据，完成模型训练，得到模型文件。
	4.4 深度学习模型推理应用	4.4.1 能够根据应用场景，加载模型，对实际数据进行模型推理。 4.4.2 能够根据应用场景，设计合理的预测服务接口方式，并完成接口开发。

工作领域	工作任务	职业技能要求
		4.4.3 能够根据应用场景，完成合适的模型转换，完成端侧部署和推理应用。

表 3 自然语言处理应用开发职业技能等级要求（高级）

工作领域	工作任务	职业技能要求
1. 语料采集与整理	1.1 语料采集	1.1.1 能够根据需要，制定采集的语料数据数量和质量要求。
	1.2 语料整理	1.2.1 能够根据实际需求，对段落、空格、字符等相关问题进行分析。 1.2.2 能够根据实际需求，制定语料整理质量要求。
2. 语料预处理	2.1 语料清洗	2.1.1 能按照模型训练要求，制定语料清洗的方式和质量要求。
	2.2 分词模型训练	2.2.1 能按照模型训练要求，制定分词的方式和质量要求。 2.2.2 能够根据实际需求，选择合适的分词工具。
	2.3 词性标注	2.3.1 能按照模型训练要求，制定词性标注的方式和质量要求。 2.3.2 能够根据实际需求，选择合适的标注方法和工具。
3. 深度学习模型训练评估	3.1 机器学习模型训练	3.1.1 完成机器学习模型训练基本方法的学习。 3.1.2 完成有监督学习模型训练方法的学习。 3.1.3 完成有监督学习模型多种评估方法的学习。
	3.2 深度学习模型认知	3.2.1 完成全连接网络模型以及原理的学习。 3.2.2 完成卷积神经网络模型的人脸识别技术的深入学习。 3.2.3 完成循环神经网络的门控算法技能的熟练使用。 3.2.4 完成长短时记忆网络的搭建实操。
	3.3 深度学习模型训练与评估	3.3.1 完成深度学习框架 TensorFlow、Mxnet 等开发环境的学习。 3.3.2 能够根据要求，完成模型训练的数据准备。 3.3.3 能够根据要求，迭代数据，完成模型训练，得到模型文件。 3.3.4 能够根据要求，加载模型，根据场景，选择合适评估方法，并进行模型的性能和效率综合评估。
	3.4 深度学习模型推理应用	3.4.1 能够根据应用场景，加载模型，对实际数据进行模型推理。 3.4.2 能够根据应用场景，设计合理的预测服务接口方式，并完成接口开发。

工作领域	工作任务	职业技能要求
		3.4.3 能够根据应用场景，完成合适的模型转换，完成端侧部署和推理应用。
	3.5 模型优化训练	3.5.1 能够适当调整模型运行参数，对模型训练效率和训练效果进行一定程度的调优。 3.5.2 能够以机器学习理论为指导，分析数据，迭代数据策略，完成特征优化，模型训练，对学习效果进行迭代。
4.自然语言处理工程应用	4.1 基础算法认知	4.1.1 完成关键词提取的解决思路和经典算法的学习。 4.1.2 完成句法分析的解决思路和经典算法的学习。 4.1.3 完成文本向量化问题的解决思路和经典算法的学习。
	4.2 垂直领域工程应用开发	4.2.1 针对用户评论、情感分析应用场景，完成工程应用方案，具有能有进行应用开发的能力。 4.2.2 针对客服提问、智能回答应用场景，了解工程应用方案，能够进行应用开发。 4.2.3 针对归纳、总结、生成文本摘要应用场景，完成工程应用方案，能够进行应用开发。 4.2.4 针对文本主题分析，自动分类应用场景，完成工程应用方案，能够进行应用开发。

## 参考文献

- [1] 国务院关于印发新一代人工智能发展规划的通知
- [2] 教育部关于印发《职业教育专业目录（2021年）》的通知（教职成〔2021〕2号）
- [3] 《教育部关于公布2020年度普通高等学校本科专业备案和审批结果的通知》（教高函〔2021〕1号）
- [4] 《教育部关于公布2019年度普通高等学校本科专业备案和审批结果的通知》（教高函〔2020〕2号）
- [5] 中等职业学校专业教学标准（2014年试行）
- [6] 高等职业学校专业教学标准（2018年）
- [7] 本科专业类教学质量国家标准（2018年）
- [8] 国家职业技能标准编制技术规程（2018年版）
- [9] 中华人民共和国职业分类大典（2015年版）
- [10] 战略性新兴产业分类（2018）
- [11] GB/T 5271.34-2006 人工智能神经网络
- [12] GB/T 1.1-2009 标准化工作导则
- [13] GB/T 5271.31-2006 人工智能机器学习