

# 高级软件工程实践课(2023) 课程要求说明



# 一、总体要求

- 综合使用先进的软件工程技术，结合经验软件工程知识，基于知识图谱技术，针对开源社区开发一个有实际应用价值的应用，能够识别出社区博客中讨论的技术，并且有效挖掘博客中与技术相关的知识，完成知识图谱构建、展示，并基于所构建图谱完成自动问答、智能推荐、智能搜索、自动实体评价等任务。

## 二、功能要求- I

- 1.知识建模
- (1)问题模型，明确使用知识图谱所解决的软件工程问题
- (2)本体模型，识别社区博客中数据在问题背景下的本体模型

为保证本体模型的建模质量，需标注2000条及以上的数据。在标注本体时，必须进行Aspect Based Sentiment Analysis (ABSA)标注，即标注文本向该本体表达的情绪，其他与技术相关的标注可自选，如技术可运用场景、技术优缺点、技术发展历程等，或其他能够表现出技术发展趋势的内容。

- (3)模型可视化展示

## 二、功能要求-2

- 2. 社区选择
- 开源社区包括但不限于以下四类：
  - (1)中文综合类社区：CSDN、思否、掘金、InfoQ
  - (2)英文综合类社区：Stack Overflow、CodeProject、Reddit、HackerNews
  - (3)企业开发者社区：华为云开发者社区、腾讯云开发者社区、阿里云开发者社区、51CTO技术网站
  - (4)专业类社区（只关注某类技术）：ChinaUnix、CNodeJS、Ruby China

## 二、功能要求-3

- 3.知识获取及构建

- (1)数据爬取，能够自动爬取社区博客文章并有效识别出其中的技术
- (2)数据规模，不少于2000条数据
- (3)识别结构化/半结构化/非结构化数据
- (4)数据存储，可以使用关系型或者非关系型数据库
- (5)图谱可视化展示

- 4.知识计算

- (1)描述所使用的知识推理方法及技术
- (2)通过知识推理发现隐含关系以及知识

## 二、功能要求-4

- 5.知识应用
  - (1)描述所解决软件工程问题
  - (2)介绍所使用的语义搜索、知识问答或推荐等方法
  - (3)展示系统
  
- 6.知识强度(可信度)评价
  - (1)对知识获取的证据来源、形式、声誉、数量等进行分类统计
  - (2)根据以上证据强度信息对知识可信度进行评价计算
  - (3)将知识强度融合进知识图谱，支持可视化和查询检索

## 二、功能要求-5

- 7.参考应用
- (1)答案生成(Community Answer generation Based on Knowledge Graph , Knowledge Graph Based Question Routing for Community Question Answering)
- (2)技术全景查询(TechLand:Assisting Technology Landscape Inquiries with Insights from Stack Overflow )
- (3)API警告增强(Improving API Caveats Accessibility by Mining API Caveats Knowledge Graph )
- (4)Thoughtworks技术雷达 (<https://www.thoughtworks.com/radar>)
- (5)自动实体评价（按照实体，识别并汇总博客中的相关文本，对实体的相关文本进行情绪分析，最后以可视化图表展示对于同一实体的评价，可参考Pattern-Based Mining of Opinions in Q&A Websites）
- (6)标注指南（SemEval 2014 Task 4 ABSA Annotation Guidelines SemEval是一个国际性的语义评估比赛，上述标注指南介绍了如何在餐厅和笔记本电脑的客户评论中标注情绪方面对（aspect terms + aspect term polarity）。该指南可用于指导我们在SE文本中进行类似的标注工作。）
- (7)标注工具(可选用)：YEDDA 轻量级标注工具 YEDDA 可用于文本跨度标注，并计算标注一致性
- (8)仅供参考，不限于以上应用

## 二、功能要求-6

- 8.相关概念

- (1)ABSA: ABSA旨在识别文本中的方面 (aspect) , 然后确定每个方面的情感极性 ( sentiment ) , 即积极、中性、 消极。这允许我们更精确地了解人们对于特定方面的情感如何, 而不仅仅是整体的情感。

- (2)ABSA in SE: ABSA in SE 允许我们深入分析软件产品的不同方面, 如性能、用户界面、可维护性等, 以了解用户的情感和意见。这有助于精准识别问题、满足用户需求、提高软件质量和用户体验。



## 三、开放服务

- 1、针对上述功能提供接口服务
- 2、服务接口满足Restful 编程风格

## 四、软件过程/技术要求

- 1.关键功能具有软件测试
- 2.实现持续集成/持续交付(含自动化测试)
- 3.基于SpringCloud框架实现微服务架构、容器化部署
- 4.具体功能实现上可采用适合的数据挖掘、机器学习等技术，但勿做过度开发

## 五、作业提交要求

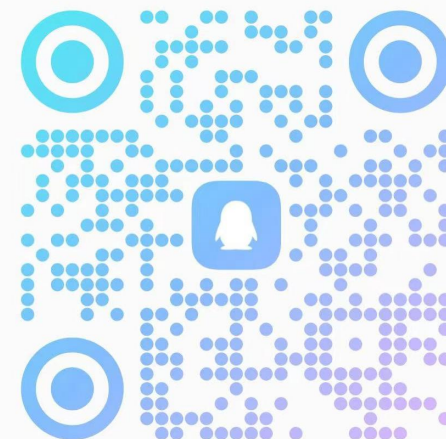
- 1.每双周日晚8点之前提交两周进展说明
  - (1)主要强调近两周的项目进展和后面两周的交付计划
  - (2)以视频形式(时长5分钟)进行整体说明，其中提到的主要制品以文档形式说明
- 2.四周后完成数据标注，八周后完成平台上线部署，并随时有可执行的发布版本(功能可持续实现)供抽检
- 3.学期末现场汇报
  - (1)20分钟PPT+视频/现场演示
  - (2)5分钟项目问题回答

# 软件定义世界 数智驱动未来



2023高级软件工...

群号: 920519985



扫一扫二维码，入群聊