

2022-2023 第一学期舆情分析期末试题（闭卷考试）

回忆版

注意事项：姬东鸿的 PPT 几乎没有考什么内容，都是李霏 PPT 上的。试卷上的字很少，即使是回忆版，内容与原卷也可以说完全一致了。

一、名词解释（每题 5 分，共 30 分）

- 1, 自然语言
- 2, 分词
- 3, 实体识别
- 4, 循环神经网络
- 5, 语言模型
- 6, 词向量

二、简答题（每题 10 分，共 40 分）

- 1, 举出三个分类任务，并说明应用场景。
- 2, 简述卷积神经网络，循环神经网络，Transformer 模型的特点（区别）
- 3, 简述 Word2Vec 训练词向量的原理和方法
- 4, 简述 Transformer 中 self-attention 的原理和计算流程

三、方案设计（每题 15 分，共 30 分）

- 1, 现在有一个 15000 条数据产品的评论数据，已经人工标注了情感正负极，请设计一个基于深度学习模型的情感分类方案，要求写出流程图、模块说明、评价指标。
- 2, 现在有 10 万篇爬取的文档（无标注），1000 条句子（使用 BIO 标注了人名、地名实体），设计一个 Transformer 模型，采用预训练和 Fine-Tune 构建一个实体识别系统，要求写出流程图、模块说明、评价指标。