

2021-2022 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, 预训练
- 2, 文本蕴含
- 3, 注意力机制
- 4, 词性标注
- 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 LSTM 和 GRU 的结构，并说明区别
- 3, 简述 CBOW 和 Skip-gram 训练词向量方式的区别。
- 4, 简述 Bert 和 GPT 的结构区别与预训练的区别。

三、方案设计（每题 15 分，共 30 分）

- 1, 现在有一个 15000 条中文新闻组成的数据，标注了其中的命名实体，但是有些实体存在重叠和嵌入，例如“武汉大学”包含“武汉”这个实体。请设计一个方案进行命名实体识别，要求写出流程图、模块说明、评价指标。
- 2, 现在有一个 15000 条中文和其英文翻译的数据集。请设计一个方案进行深度学习的机器翻译任务，要求写出流程图、模块说明、评价指标。