

自然语言处理

第1讲：课程介绍

刘洋



课程基本信息

- 课程名称：自然语言处理
- 课程编号：081203M04002H
- 课程属性：专业核心课
- 课程学时：60
- 课程学分：3
- 考核方式：文献综述+编程作业+闭卷考试
- 上课时间：每周四18:10-20:50
- 上课地点：教1-101

教师与助教

教师

刘洋，教授， liuyang2011@tsinghua.edu.cn

助教

邵晨泽，研究生， shaochenze18z@ict.ac.cn

助教

刘鹏，研究生， liupeng19@mails.ucas.ac.cn

上课计划

周次	时间	内容
2	2020-09-17 18:10-20:50	第1讲：课程介绍（1学时） 第2讲：自然语言处理概述（2学时）
3	2020-09-24 18:10-20:50	第3讲：数学基础（3学时）
4	2020-10-01 18:10-20:50	放假
5	2020-10-10 18:10-20:50	第4讲：隐马尔科夫模型（2学时） 第5讲：支持向量机（1学时）
6	2020-10-15 18:10-20:50	第6讲：对数线性模型（2学时） 第7讲：神经网络（1学时）

上课计划

周次	日期	内容
7	2020-10-22 18:10-20:50	第7讲：神经网络（3学时）
8	2020-10-29 18:10-20:50	第8讲：文本分类与聚类（1学时） 第9讲：语言模型（2学时）
9	2020-11-05 18:10-20:50	第10讲：词法分析（1学时） 第11讲：句法分析（2学时）
10	2020-11-12 18:10-20:50	第11讲：句法分析（3学时）
11	2020-11-19 18:10-20:50	第12讲：语义分析（2学时） 第13讲：篇章分析（1学时）

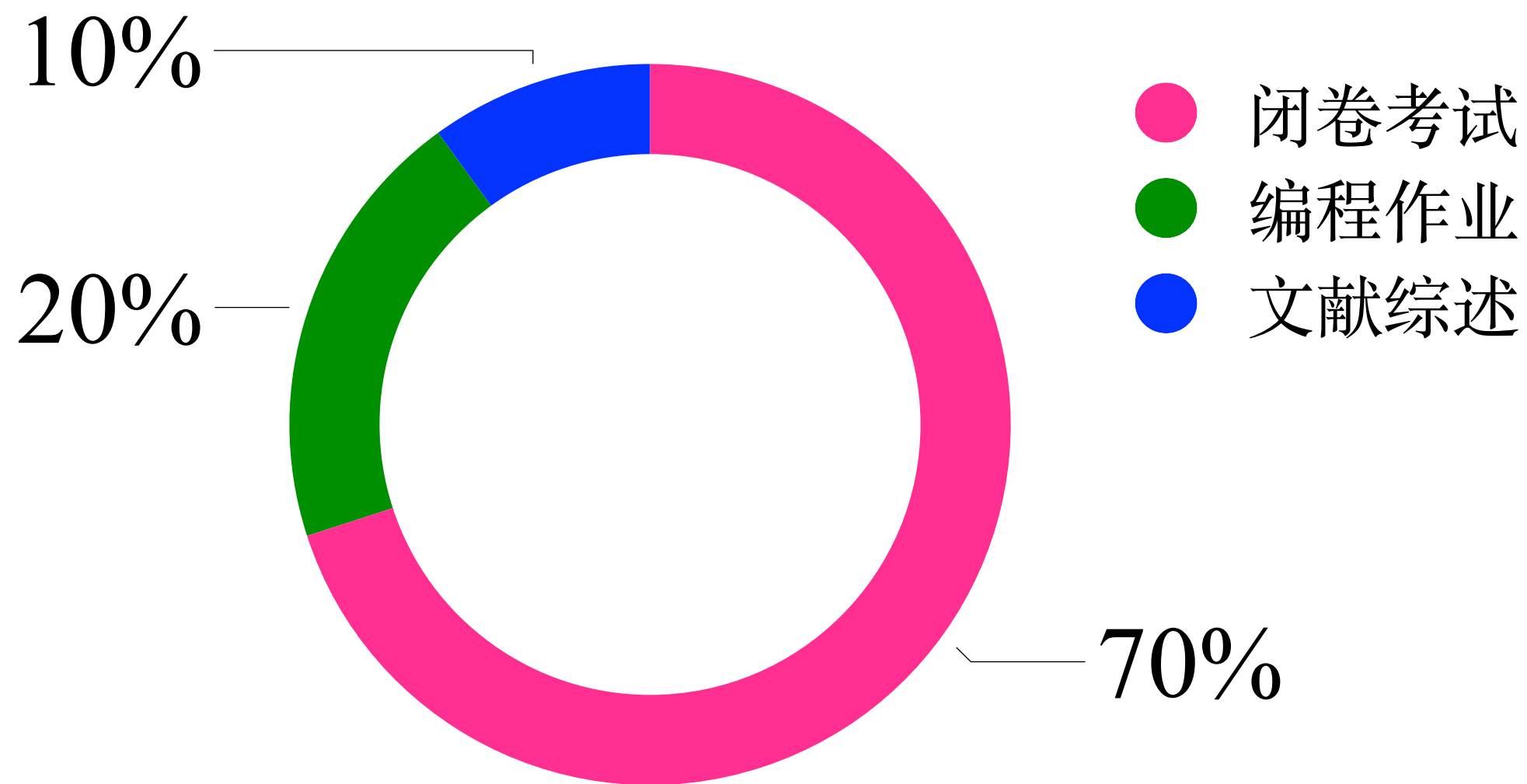
上课计划

周次	日期	内容
11	2020-11-22 08:30-11:20	第14讲：机器翻译（3学时）
12	2020-11-26 18:10-20:50	第14讲：机器翻译（3学时）
13	2020-12-03 18:10-20:50	第15讲：文本生成（2学时） 第16讲：情感分析（1学时）
14	2020-12-10 18:10-20:50	第16讲：情感分析（1学时） 第17讲：信息抽取（2学时）
14	2020-12-13 08:30-11:20	第18讲：知识图谱（3学时）

上课计划

周次	日期	内容
15	2020-12-17 18:10-20:50	第19讲：问答系统（2学时） 第20讲：对话系统（1学时）
16	2020-12-24 18:10-20:50	第21讲：多模态自然语言处理（1学时） 第22讲：预训练语言模型（2学时）
16	2020-12-27 08:30-11:20	课堂展示
17	2020-12-31 18:10-20:50	期末考试

成绩计算方法



文献综述

- 考核目标：锻炼学生查阅文献和论文写作能力，让学生对某个具体的自然语言处理方向具备较深入的了解。
- 可选领域：语言模型、词法分析、句法分析、文本分类、情感分析、文本摘要、阅读理解、诗词生成、图像标题生成、其他
- 提交形式：文献综述论文的PDF文件
- 评分标准：
 - 写作质量（50%）：篇幅合适、文字流畅、行文严谨、表示准确、排版美观
 - 学术质量（50%）：覆盖面全、追踪前沿、分类合适、评述到位、独抒己见

文献综述

- 具体要求如下：
 - 每个人独立完成；
 - 严禁抄袭现有论文，违者记0分；
 - 严禁抄袭其他同学的论文，违者抄袭者和被抄袭者均记0分；
 - 必须使用LaTeX撰写，中英文均可；
 - 论文篇幅至少6页（不含参考文献）；
 - 建议论文结构组织如下：标题、作者信息、摘要、问题定义、发展历史与主要方法分类、未来方向；
 - 选择“其他”类别必须与自然语言处理密切相关，需提交助教审核。

编程作业

- 考核目标：锻炼学生编程解决自然语言处理问题的能力，加深学生对课堂讲授内容的理解。
- 可选领域：语言模型、词法分析、句法分析、文本分类、情感分析、文本摘要、阅读理解、诗词生成、图像标题生成、其他
- 提交形式：源代码、网页端Demo、运行支撑文件、文档
- 评分标准：
 - 投入（25%）：是否主要是独立完成的代码？
 - 实验（25%）：相对于基准系统有多大提升？
 - 演示（25%）：界面的设计是否对用户友好？
 - 文档（25%）：文档是否条理清晰撰写规范？

编程作业

- 具体要求：
 - 按照小组提交作业，每个小组不超过5人；
 - 每个方向最多可接受15个小组申报，先到先得；
 - 选择“其他”类别必须与自然语言处理密切相关，需提交助教审核；
 - 可以使用任意的编程语言；
 - 每个小组自行决定实验使用的数据集和基准系统；
 - 文档的格式是通过LaTeX生成的PDF文件，主要内容包括小组成员信息及分工、任务定义、方法描述、系统框架、实现细节、数据集、基准系统、实验设置与结果、实验分析；

编程作业

- 具体要求：
 - 网页端Demo要求可以在本地或者服务器端运行，接受用户输入，返回系统输出；
 - 禁止使用网上现成的代码以很小的代价完成作业；
 - 建议文献综述与编程作业的方向一致；
 - 允许使用参加其他科创比赛的程序，但需与自然语言处理密切相关且明确表明本人贡献，需助教审核。

闭卷考试

- 考核目标：考察学生对自然语言处理重要知识点的掌握程度。
- 考试时间：预计在2020年12月31日
- 考试形式：闭卷
- 主要题型：判断题、填空题、计算题

时间表

日期	内容
2020-09-17	小组报名开始
2020-09-30	小组报名截止
2020-12-13	文献综述提交截止
2020-12-13	编程作业提交截止
2020-12-27	课堂展示（文献综述+编程作业）
2020-12-31	闭卷考试

小组报名

- 请自由组队（不超过5人），于2020年9月30日23:59之前将以下信息发给助教邵晨泽（shaochenze18z@ict.ac.cn）
 - 小组名称：取一个反映小组特点的名字，如“雁栖大侠”；
 - 小组成员：列出每位组员的姓名、学号、邮箱、手机号；
 - 所选任务：从给定的10个任务中选取（每个任务最多15组，先到先得）；
 - 备注说明：如果选择的是“其他”，请说明理由，由助教审批。

谢谢