

前周工作报告

九月

第五周

2017-09-25

2017-10-01

1. 完善前端标注 Remarkable

本周主要致力于实现前端标注页面，在完成后端逻辑，题目的生成、分流、存储、上传等API后，主动前去协助前端的实现。

- [Issue #26](#)

Some optional improvements in causality_marking

前端于测试环境上线时，考虑到的几个问题，经过讨论提出issue，并逐一解决。

- ☒ 界面中增加sid
- ☒ 刷新换题的问题
- ☒ 为问题增加Tags与status的标记
- ☒ 无题目时的处理
- ☒ 前向审阅

- [Issue #29](#)

Return indexes of selected words in answer.

返回数据的格式，与先前讨论的略有出入，强调细节后进行修改。

- ☒ answer由list转化为dict，使用Name作为key
- ☒ 确立以分词为最小粒度，C、R、Link、Range均为分词下标表示

- [Issue #30](#)

Possible empty/incorrect submission

在标记过程中发现的种种问题，写为issues以供节后更新

- ☐ 切换问题后，tag不初始化的问题
- ☐ 多次点击可能因为延迟造成大量过题的现象

2. 完善后端 Autodoc

继续为前端完善所需的后端API，并在标注过程中持续为后端补充连接词表 `link_word` 和关联词库 `link_couple`，相关代码可详见 [【LinkCatch】](#)

- [Add data_post for test environment \(http://tl.bondowner.cn\)](http://tl.bondowner.cn).

实现通过HTTP协议上传题目的API

- [Change the way of output for Py3 adaptation.](#)
更改输出方式以同时适配Python2和Python3
- [Completely replaced original link_extractor with new class.](#)
将新的LinkCatch类最后几个函数实现，试图完全替代原先的LinkSplit类
- [Add API for preview all o301 data.](#)
写了一个脚本用以统计总计题目数量及单关联词的题目数量
- [packages will be import only when used in utils.](#)
更新工具类utils，每当使用某个函数时才进行import，减少初始化时所消耗的时间
- [Change checking_dumps for Permission denied.](#)
考虑到文件读写时可能会出现权限不足的问题导致程序崩溃，更新了权限检测，增强程序健壮性
- [batch dump for preview question_data.](#)
成功上线，批量为前端post题目的同时，dump为本地文件以供检查
- [update original cs_extractor for key_error.](#)
修改字典获取方式，从 `dict[key]` 更改为 `dict.get(key)` 以防止出现KeyError的问题
- [noting post_url for test_env and formal_env.](#)
- [Question pushing had been shunted into two ways](#)
新增**分流**机制，即每次抽取出题目后先让规则库进行判别，简单判别的题目直接生成该题目的答案存于本地，不进入前端的题库中，当前仅实现一种规则用于测试：“同segment内的 **前置词** + **所致**”(这种规则出现相当频繁，直接解决了约一万道题目的标注)
 - [update shunt rules with link_couple.](#)
 - [change strategy of link_couple extracting.](#)
 - 当前需要人工标注数：21808 / 32095（自动标注规则数：1）
- 修复调试与展示过程中出现的小问题
 - [decrease log text during posting.](#)
 - [fix doc_idx type error](#)

3. 继续更新规则库、正则与词表

- [Update link_words, modify marking rules.](#)
- [Update link_list.](#)
- [Update link_list.](#)
- [Add new special cases.](#)
- [Update link_word list](#)
- [update Regex for link_couple with comma](#)
- [Update special cases and link_word list during marking.](#)
- [Update link_list.](#)
- [Update link_list.](#)
- [noting new possible link_couple examples.](#)

Other Works

Teaching assistant

继续于玉泉路担任助教职责

本周没有作业批改任务，下周为期中考试。

Pytorch Research

调研中：Torch Self-Learning

[Torch_Basic_Learning](#)

[Torch_NN_Learning](#)

PPT Modify

PPT for AI Course

[Finished]

将最后一个文件夹中的PPT过了一遍，需要修改的地方仅有一处拼写错误和一处格式微调，至此全四个文件夹的更新全部结束，如果有需要修改的可以继续随时修改。

Website Setup

初期使用以下定义：

[aifundamental.com](#) for 国内访问（oschina源）

[ai-fundamental.com](#) for 国际访问（github源）

[pages assets](#) 作为矢量图等图像存放方式

[github Release](#) 作为课件等在线存储载体，后期采用类似斯坦福大学的课程网页形式

owner权限邮箱均设置为 luop@ict.ac.cn

辅助修改权限设置为 okcd00@qq.com

Weekly Timetable

- [2017-10-07](#) 周六：到北京后到计算所的办公室报到

Notes

- Pytorch learning.
- Marking for Remarkable