

机器阅读理解

挑战当前优秀的阅读理解系统尚不能正确回答的问题，
全面评测机器进行深度语言理解以回答复杂问题的能力。

报名成功

比赛介绍

数据下载

结果提交

获奖名单&排行榜

新闻中心

国内用户下载

<div>train_raw.zip</div> <div>训练集</div> <div>https://nlpc-du.cdn.bcebos.com/reading/train_raw.zip</div>	<div>dev_raw.zip</div> <div>开发集</div> <div>https://nlpc-du.cdn.bcebos.com/reading/dev_raw.zip</div>	<div>test1_raw.zip</div> <div>测试集1</div> <div>https://nlpc-du.cdn.bcebos.com/reading/test1_raw.zip</div>	<div>test2_raw.zip</div> <div>测试集2</div> <div>5月13日开放下载</div>
<div>train_preprocessed.zip</div> <div>训练集</div> <div>https://nlpc-du.cdn.bcebos.com/reading/train_preprocessed.zip</div>	<div>dev_preprocessed.zip</div> <div>开发集</div> <div>https://nlpc-du.cdn.bcebos.com/reading/dev_preprocessed.zip</div>	<div>test1_preprocessed.zip</div> <div>测试集1</div> <div>https://nlpc-du.cdn.bcebos.com/reading/test1_preprocessed.zip</div>	<div>test2_preprocessed.zip</div> <div>测试集2</div> <div>5月13日开放下载</div>

海外用户下载

<div>train_raw.zip</div> <div>训练集</div> <div></div>	<div>dev_raw.zip</div> <div>开发集</div> <div></div>	<div>test1_raw.zip</div> <div>测试集1</div> <div></div>	<div>test2_raw.zip</div> <div>测试集2</div> <div>5月13日开放下载</div>
<div>train_preprocessed.zip</div> <div>训练集</div> <div></div>	<div>dev_preprocessed.zip</div> <div>开发集</div> <div></div>	<div>test1_preprocessed.zip</div> <div>测试集1</div> <div></div>	<div>test2_preprocessed.zip</div> <div>测试集2</div> <div>5月13日开放下载</div>

数据介绍

2. 开发集：共2239个问题，包含答案，用于竞赛模型训练和参数调试。
3. 测试集1：共3398个问题，不包含答案，用于参赛者在平台上自助验证模型效果。为了防止针对测试集的调试，数据中将会额外加入混淆数据。
4. 测试集2：是本次竞赛最终测试数据，共6851个问题，其中包含测试集1，不包含答案。为了防止针对测试集的调试，数据中将会额外加入混淆数据。该部分数据结果不能在平台上自助验证。

数据样本

平台提供的数据为JSON文件格式，样例如下：

```
{
  "question_id": 186358,
  "question_type": "YES_NO",
  "question": "上海迪士尼可以带吃的进去吗",
  "documents": [
    {
      'paragraphs': ["text paragraph 1", "text paragraph 2"]
    },
    ...
  ],
  "answers": [
    "完全密封的可以，其它不可以。", // answer1
    "可以的，不限量的。只要不是易燃易爆的危险物品，一般都可以带进去的。", //answer2
    "罐装婴儿食品、包装完好的果汁、水等饮料及包装完好的食物都可以带进乐园，",
    "但游客自己在家制作的食品是不能入园，因为自制食品有一定的安全隐患。" // answer3
  ],
  "yesno_answers": [
    "Depends", // corresponding to answer 1
    "Yes", // corresponding to answer 2
    "Depends" // corresponding to answer 3
  ]
}
```

*更多样例和详细数据格式说明参见数据集中包含的“数据格式说明”文档。

入门参考

数据集[论文](#): DuReader: a Chinese Machine Reading Comprehension Dataset from Real-world Applications

[基线系统](#): 两个开源的阅读理解基线系统，及结果评价程序源码

[百度AI数据集开放平台](#): 自行下载已发布的阅读理解数据集，在线进行系统测试

联系我们

任何与本次技术竞赛相关的问题，请随时联系竞赛会务组。

竞赛会务组邮箱：lic2019@126.com