比赛链接位置 https://www.statmt.org/wmt20/quality-estimation-task.html
点击 result 可以看到结果，点击 submissions 得到当年比赛的各个队伍的提交文件

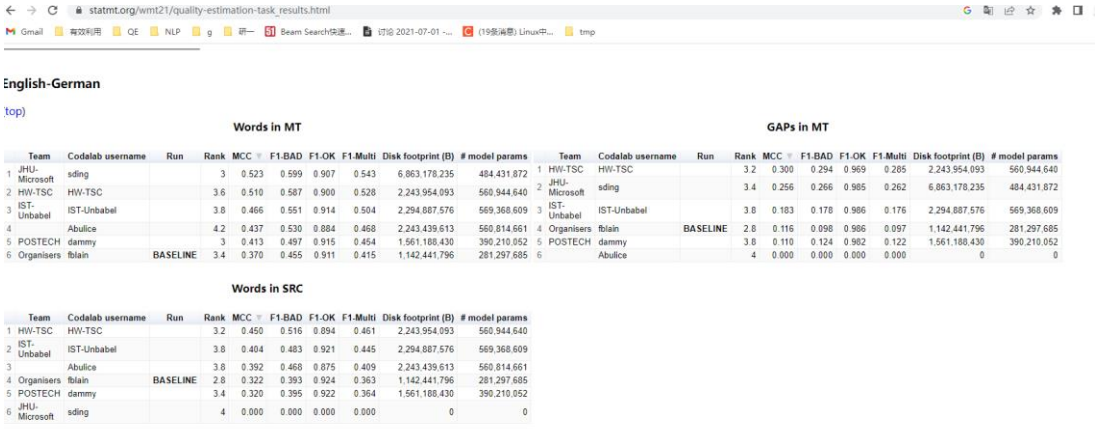
以 wmt20 ende qe 任务为例
打开 team NJUNLP 的提交文件，里面是 predictions_mt.txt 文件
预测的内容是 gap mt gap mt gap …的形式

wmt20 官方这个结果是 mt 和 gap 合在一起算的结果

English-Chinese

	Team	Codalab username	Run	Target MCC	Target F1-BAD	Target F1-OK	Source MCC	Source F1-BAD	Source F1-OK
1	NiuTrans	NiuTrans		0.610	0.723	0.887	0.308	0.666	0.639
2	HW-TSC	yuriak		0.587	0.714	0.866	0.000	0.000	0.000
3	NICT Kyoto	Raphael_NICT		0.582	0.704	0.878	0.336	0.668	0.669
4	IST and Unbabel	ddk		0.575	0.706	0.850	0.287	0.705	0.410
5	IST and Unbabel	Joao1996		0.567	0.701	0.842	0.287	0.705	0.403
6	NJUNLP	cuiqu		0.551	0.672	0.877	0.000	0.000	0.000
7	Organisers	erickrf	BASELINE	0.509	0.658	0.849	0.270	0.682	0.547

下图是 wmt21 的最终结果，可以看到 mt gap src 三个指标是分开算的，也就是每年官方的计算方法可能有变动



回到 wmt20enzh，njunlp 提交的结果是 predicions_mt.txt
对应的正确的结果是 all.tags 文档，压缩包里有
（这个 all.tags 文档是我们自己算的，不是当时官方公开的，当时官方误传了部分测试集文件，随后官方很快删了，但被之前的师兄下载下来了。我们然后通过这个官方上传的文件计算出了 all.tags 文件，所以这个并不是完全的标准答案，和正确答案可能有一点点偏差）

比较提交结果 predicions_mt.txt 和 all.tags 文档
用的 python 文件 f1cal_mcc.py，注意更改存的文档的位置，运行这个 python 文件

```
def my_main():

    system_file = "/home/zhangy/predictions_mt.txt"
    gold_file = "/home/zhangy/all.tags"
```

得到的结果是

```
[0.8762554775662156, 0.6701126847927221]
0.5871899106362256
0.548433823258102
finished
```

第一行是 f1ok, f1bad
第二行是 f1-mult=f1ok* f1bad
第三行是 mcc

下图是网页上官方计算的最终结果，和我们计算的基本一致，有一点误差，都在 0.003 内
这是因为用于计算的 all.tags 是我们自己算的，不完全准确

6	NJUNLP	cuiqu		0.551	0.672	0.877