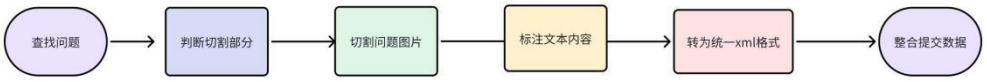
大学拍照搜题数据标注手册

一、标注目标：

将教辅材料与试卷每一页中的题目切割下来并转写，形成结构化数据，以供图像搜索使用。

二、大学拍照搜题项目工作流程：



三、判断切题

切割教辅材料与试卷的每一页中的题目

（1）题目含义：

要求回答或解释的内容，实际操作过程中需要将文字题干及问题、配图全部切割下来。若题目中包含答案，需要将答案抹除

（2）题目主要包括几大类型及举例：

下举例均为一个切割部分>

a.解答题：大题单问，需要在答题空白处撰写文字、数字答案的题目。

例如：高等数学试卷 18—19 年

三、计算题（本题共 8 小题，每小题 6 分，共 48 分）

1. $\lim\limits_{n\to\infty}(\sqrt{n+\sqrt{n}}-\sqrt{n})$

b.选择题：同时拥有多个不同选项并需要选出正确答案的题目。

例如：2023 年全国硕士研究生统一考试数学（一）

（1）曲线 $y=x\ln(e+\frac{1}{x-1})$ 的斜渐近线方程为（　　）

（A）$y=x+e$　（B）$y=x+\frac{1}{e}$　（C）$y=x$　（D）$y=x-\frac{1}{e}$

c.填空题：需要填写答案使题目完整的题目。

例如：2023 年全国硕士研究生统一考试数学（一）

（14）设连续函数 $f(x)$ 满足：$f(x+2)-f(x)=x$，$\int_0^2 f(x)dx=0$，则 $\int_1^3 f(x)dx=$_____

d.判断题：需要判断题干是否符合知识点的题目。

例如：

6．算法可以用不同的语言描述，如果用 C 语言或 PASCAL 语言等高级语言来描述，则算法实际上就是程序了。（ ）　【西安交通大学 1996　二、7（3 分）】

e.画图题：需要补充图画答案的题目。

例如：CAD 绘图题

练习

（1）利用正交与对象追踪绘制图 3.1 与图 2.1 右侧所示图形;

（2）利用正交与对象追踪绘制图 4.1 所示图形;

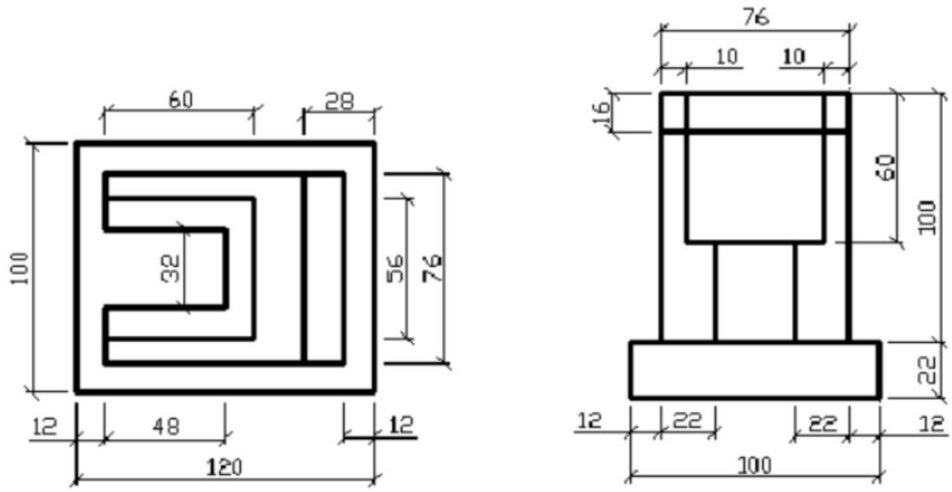


图 4.1　U 形桥台投影图`

f.证明题（应用题）：大题多问（多行），排版的时候有着留白以回答大量 ans 内容的题目。

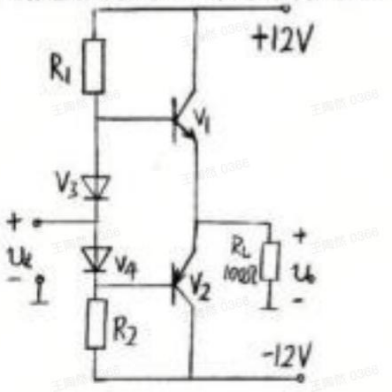
例如：物理应用题

4、电路如图所示，设 $U_{CES}=0$ 试回答下列问题：（6 分）

（1）ui=0 时，流过 RL 的电流有多大?

（2）若 V3、V4 中有一个接反，会出现什么后果?

（3）为保证输出波形不失真，输入信号 ui 的最大幅度为多少? 管耗为多少?

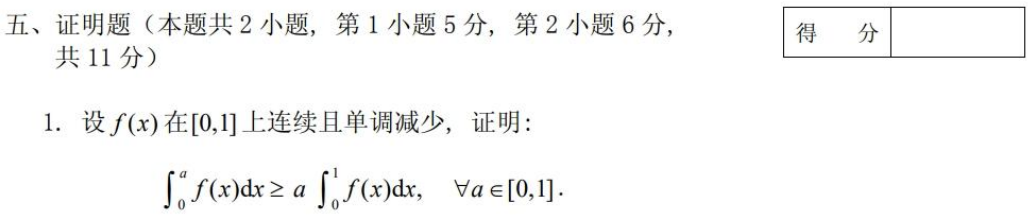


四、切图的具体说明：

以标注文件页数中的主题干（没有主题干，问题不成立）来判定切割数量，一个主题干为一个切割部分，如遇到跨页，就有几页切几个图，不需要切图之后拼在一起，例如。

（1）数学计算题的切图方式

如图所示主题干为 2 个主题干，因此切割为 2 个部分，以进行下一步标注工作。例如下图，



五、证明题（本题共 2 小题，第 1 小题 5 分，第 2 小题 6 分，共 11 分）

得　分	

1. 设 $f(x)$ 在 $[0,1]$ 上连续且单调减少，证明：

$$\int_0^a f(x)\mathrm{d}x \geq a\int_0^1 f(x)\mathrm{d}x, \quad \forall a \in [0,1].$$

2. 设 $f(x)$ 在 $[0,1]$ 上二阶可导，且 $f(0)=\frac{1}{2}$, $f(\frac{1}{2})=1$, $f(1)=0$. 证明：

（1）$\exists \zeta \in (0,1)$，使得 $f'(\zeta)=0$.（4 分）

（2）$\exists \eta \in (0,1)$，使得 $f''(\eta)(1-\eta)-2f'(\eta)=0$.（2 分）

（2）选择题的切图方式

如图所示主题干为 6 个主题干，因此切割为 6 个部分，以进行下一步标注工作。例如下图，

《机械制造基础》期末复习题答案

一、　单项选择题（每小题　1 分，共　30 分）

在下列每小题的四个备选答案中选出一个正确的答案，并将其字母标号填入题干的括号内。

1、以下哪种加工，前角应选大些。（　C　）

A. 加工脆性材料　B. 工件材料硬度高；

C. 加工塑性材料；　D.　脆性或塑性材料。

2、以下哪种磨削，选用软的砂轮。（　A　）

A. 磨削硬材料；　　B. 磨削软材料；

C. 磨削断续表面；　　D. 精磨。

3、粗加工选择切削用量的基本原则是（　A　）。

A. 选取大的切削深度；　B. 在单位时间内切除尽量少的加工余量；

C. 选取小的切削深度；　D. 选取大的切削速度。

4、精加工时，应选用哪种切削液。（　C　）

A. 水溶液；　B. 乳化液；

C. 切削油；　D. 温度较高的水溶液。

5、车床刀架的横向运动方向与车床回转轴线不垂直，车出的工件将呈现出（　A　）。

A. 端面中凸形；　　B. 腰鼓形；

C. 锥度；　　D. 腰鼓形和锥度。

6、钻孔有两种基本方式，其一是钻头不转，工件转，这种加工方式容易产生（　B　）误差。

A. 轴线歪斜；　　B. 锥度；

C. 轴线歪斜和锥度；　　D 轴线歪斜和腰鼓形。

（3）数学多题切图方式

如图所示本页一共有 3 个不同的主提干，因此切割为三个部分，以进行下一步标注工作。例如下图，

8. 设 $V=\mathbb{R}^3=\mathbb{R}e_1+\mathbb{R}e_2+\mathbb{R}e_3$, $V_1=\mathbb{R}e_1$, $V_2=\mathbb{R}e_1+\mathbb{R}e_2$, $V_3=\mathbb{R}e_1+\mathbb{R}(e_2+e_3)$, 其中 $e_1=(1,0,0)^T$, $e_2=(0,1,0)^T$, $e_3=(0,0,1)^T$ 是 $\mathbb{R}^3$ 的自然基. 定义线性映射

$$\psi_1: V\to \mathbb{R}e_2+\mathbb{R}e_3,$$
$$xe_1+ye_2+ze_3\to \psi_1(xe_1+ye_2+ze_3)=ye_2+ze_3 \quad \text{对所有的 } x,y,z\in\mathbb{R};$$
$$\psi_2: V\to \mathbb{R}e_3,$$
$$xe_1+ye_2+ze_3\to \psi_2(xe_1+ye_2+ze_3)=ze_3 \quad \text{对所有的 } x,y,z\in\mathbb{R};$$
$$\psi_3: V\to \mathbb{R}(e_3-e_2),$$
$$xe_1+ye_2+ze_3\to \psi_3(xe_1+ye_2+ze_3)=(y-z)(e_3-e_2) \quad \text{对所有的 } x,y,z\in\mathbb{R}.$$

(a) 给出 ψ_1 的核和 ψ_1 在基 e_1,e_2,e_3 下的矩阵. 进一步证明: $V/V_1\approx\mathbb{R}e_2+\mathbb{R}e_3$ 且 $\mathbf{0}\to V_1\to V\to \mathbb{R}e_2+\mathbb{R}e_3\to\mathbf{0}$ 是正合的.

(b) 给出 ψ_2 的核和 ψ_2 在基 e_1,e_2,e_3 下的矩阵. 进一步证明: $V/V_2\approx\mathbb{R}e_3$ 且 $\mathbf{0}\to V_2\to V\to \mathbb{R}e_3\to\mathbf{0}$ 是正合的.

(c) 给出 ψ_3 的核和 ψ_3 在基 e_1,e_2,e_3 下的矩阵. 进一步证明: $V/V_3\approx\mathbb{R}(e_3-e_2)$ 且 $\mathbf{0}\to V_3\to V\to \mathbb{R}(e_3-e_2)\to\mathbf{0}$ 是正合的. 最后, 画图表示商空间 V/V_3 的同构空间.

9. (a) $0\xrightarrow{g} V\xrightarrow{f} W$ 是正合的 $\Longleftrightarrow$ f 是单射. (序列正合 $\Rightarrow 0=\mathrm{Img}(g)=\mathrm{Ker}(f)\Rightarrow f$ 是单射. 反之亦然.)

(b) $V\xrightarrow{f} W\xrightarrow{g}\mathbf{0}$ 是正合的 $\Longleftrightarrow$ f 是满射. (序列正合 $\Rightarrow \mathrm{Img}(f)=\mathrm{Ker}(g)=W\Rightarrow f$ 是满射. 反之亦然.)

(c) $\mathbf{0}\to V\to W\to Z\to\mathbf{0}$ 是正合的 $\Longleftrightarrow$ V 同构于 W 的一个子空间且 Z 同构于商空间 W/V.

10. 设 V 是一个 F 向量空间, U, W 为 V 的子空间. 假设 $U\subseteq W$.

(a) 证明映射 p:

$$V/U\to V/W,$$
$$v+U\mapsto v+W$$

是一个定义明确的线性映射, 即证明: 如果

$$v_1+U=v_2+U,$$

则

$$p(v_1+U)=p(v_2+U).$$

(b) 映射 p 的核为

$$\mathrm{Ker}\,p:=\{v+U: p(v+U)=0+W\}.$$

证明: $\mathrm{Ker}\,p=W/U$.

(c) 证明: 如下序列

$$0\to W/U\to V/U\xrightarrow{p} V/W\to 0$$

是正合的.

（4）文字类的切图方式

以英语阅读题为例，阅读原文为大题干，要根据大题干回答的小题目均为小题干，大题干+小题干切割为一道题（一条数据），以进行下一步的标注工作。例如下图的完形填空，完型填空的英语原文为大题干，根据此原文需要答题的 1-20 为小题干，大题干+小题干为一条数据。

Section I　Use of English

Directions:

Read the following text. Choose the best word(s) for each numbered blank and mark A, B, C or D on the ANSWER SHEET. (10 points)

People have speculated for centuries about a future without work. Today is no different, with academics, writers, and activists once again __1__ that technology is replacing human workers. Some imagine that the coming work-free world will be defined by __2__: A few wealthy people will own all the capital, and the masses will struggle in an impoverished wasteland.

A different and not mutually exclusive __3__ holds that the future will be a wasteland of a different sort, one __4__ by purposelessness: Without jobs to give their lives __5__, people will simply become lazy and depressed. __6__, today's unemployed don't seem to be having a great time. One Gallup poll found that 20 percent of Americans who have been unemployed for at least a year report having depression, double the rate for __7__ Americans. Also, some research suggests that the __8__ for rising rates of mortality, mental-health problems, and addiction __9__ poorly-educated, middle-aged people is a shortage of well-paid jobs. Perhaps this is why many __10__ the agonizing dullness of a jobless future.

But it doesn't __11__ follow from findings like these that a world without work would be filled with unease. Such visions are based on the __12__ of being unemployed in a society built on the concept of employment. In the __13__ of work, a society designed with other ends in mind could __14__ strikingly different circumstances for the future of labor and leisure. Today, the __15__ of work may be a bit overblown. "Many jobs are boring, degrading, unhealthy, and a waste of human potential," says John Danaher, a lecturer at the National University of Ireland in Galway.

These days, because leisure time is relatively __16__ for most workers, people use their free time to counterbalance the intellectual and emotional __17__ of their jobs. "When I come home from a hard day's work, I often feel __18__," Danaher says, adding, "In a world in which I don't have to work, I might feel rather different" – perhaps different enough to throw himself __19__ a hobby or a passion project with the intensity usually reserved for __20__ matters.

1. [A] boasting	[B] denying	[C] warning	[D] ensuring
2. [A] inequality	[B] instability	[C] unreliability	[D] uncertainty
3. [A] policy	[B] guideline	[C] resolution	[D] prediction
4. [A] characterized	[B] divided	[C] balanced	[D] measured
5. [A] wisdom	[B] meaning	[C] glory	[D] freedom
6. [A] Instead	[B] Indeed	[C] Thus	[D] Nevertheless
7. [A] rich	[B] urban	[C] working	[D] educated
8. [A] explanation	[B] requirement	[C] compensation	[D] substitute
9. [A] under	[B] beyond	[C] alongside	[D] among
10. [A] leave behind	[B] make up	[C] worry about	[D] set aside
11. [A] statistically	[B] occasionally	[C] necessarily	[D] economically
12. [A] chances	[B] downsides	[C] benefits	[D] principles
13. [A] absence	[B] height	[C] face	[D] course
14. [A] disturb	[B] restore	[C] exclude	[D] yield
15. [A] model	[B] practice	[C] virtue	[D] hardship
16. [A] tricky	[B] lengthy	[C] mysterious	[D] scarce
17. [A] demands	[B] standards	[C] qualities	[D] threats
18. [A] ignored	[B] tired	[C] confused	[D] starved
19. [A] off	[B] against	[C] behind	[D] into
20. [A] technological	[B] professional	[C] educational	[D] interpersonal

（5）判断题的切图方式

如图所示本页一共有 1 个大题干，6 个小题干。因此切割为 1 个部分，以进行下一步标注工作。

23. 音位是：

①按语音的社会属性划分出来的。（　　）

②按语音的物理性质划分出来的。（　　）

③按语音的生理性质划分出来的。（　　）

④按语音的物理性质和生理性质划分出来的。（　　）

⑤一个语音系统中能够区别意义的最小语音单位。

（　　）

⑥ 个语音系统中的最小语音单位。（　　）

（6）大小题的切图方式

所有 quiz 在标注时都需要分别切图，并标注上 quiz 图片来源。

a.物理应用题的切图方式

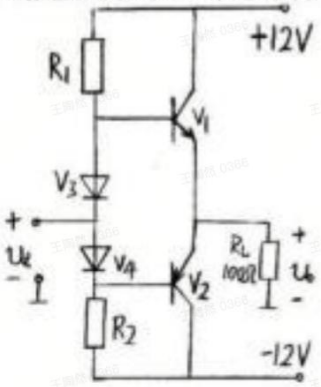
题干加题干配图切一个图片、题干配图再单独切一个图片，例如以下图片，

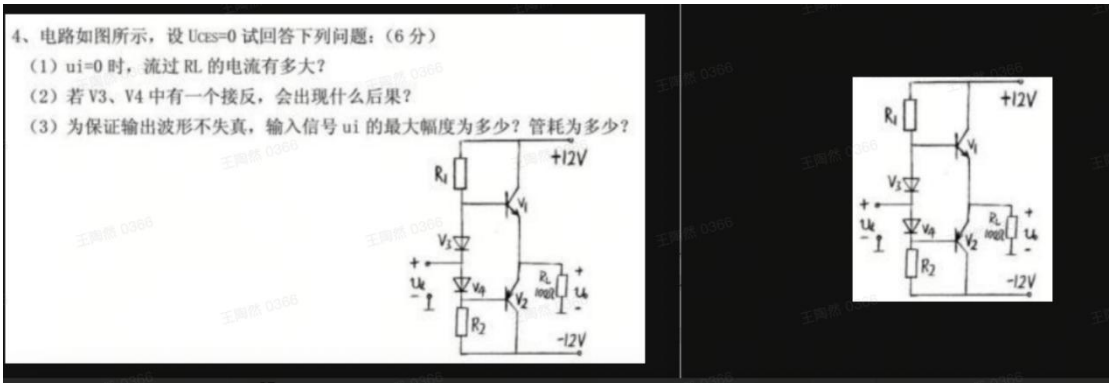
4、电路如图所示，设 $U_{CES}=0$ 试回答下列问题：（6 分）

（1）ui=0 时，流过 RL 的电流有多大？

（2）若 V3、V4 中有一个接反，会出现什么后果？

（3）为保证输出波形不失真，输入信号 ui 的最大幅度为多少？管耗为多少？





b.英语阅读题的切图方式

切英语题时，一道英语题有几页就切几页，不需要额外拼图为一张。

假如单道英语阅读题如果有两页，那么就切两个图，例如：

Text 2

Grade inflation – the gradual increase in average GPAs (grade-point averages) over the past few decades – is often considered a product of a consumer era in higher education, in which students are treated like customers to be pleased. But another, related force – a policy often buried deep in course catalogs called "grade forgiveness" – is helping raise GPAs.

Grade forgiveness allows students to retake a course in which they received a low grade, and the most recent grade or the highest grade is the only one that counts in calculating a student's overall GPA.

The use of this little-known practice has accelerated in recent years, as colleges continue to do their utmost to keep students in school (and paying tuition) and improve their graduation rates. When this practice first started decades ago, it was usually limited to freshmen, to give them a second chance to take a class in their first year if they struggled in their transition to college-level courses. But now most colleges, save for many selective campuses, allow all undergraduates, and even graduate students, to get their low grades forgiven.

College officials tend to emphasize that the goal of grade forgiveness is less about the grade itself and more about encouraging students to retake courses critical to their degree program and graduation without incurring a big penalty. "Ultimately," said Jack Miner, Ohio State University's registrar, "we see students achieve more success because they retake a course and do better in subsequent courses or master the content that allows them to graduate on time."

That said, there is a way in which grade forgiveness satisfies colleges' own needs as well. For public institutions, state funds are sometimes tied partly to their success on metrics such as graduation rates and student retention – so better grades can, by boosting figures like those, mean more money. And anything that raises GPAs will likely make students – who, at the end of the day, are paying the bill – feel they've gotten a better value for their tuition dollars, which is another big concern for colleges.

Indeed, grade forgiveness is just another way that universities are responding to consumers' expectations for higher education. Since students and parents expect a college degree to lead to a job, it is in the best interest of a school to turn out graduates who are as qualified as possible – or at least appear to be. On this, students' and colleges' incentives seem to be aligned.

26. What is commonly regarded as the cause of grade inflation?

[A] The change of course catalogs.

[B] Students' indifference to GPAs.

[C] Colleges' neglect of GPAs.

[D] The influence of consumer culture.

27. What was the original purpose of grade forgiveness?

[A] To help freshmen adapt to college learning.

[B] To maintain colleges' graduation rates.

[C] To prepare graduates for a challenging future.

[D] To increase universities' income from tuition.

28. According to Paragraph 5, grade forgiveness enables colleges to

[A] obtain more financial support.

[B] boost their student enrollments.

[C] improve their teaching quality.

[D] meet local governments' needs.

29. What does the phrase "to be aligned" (Line 5, Para. 6) most probably mean?

[A] To counterbalance each other.

[B] To complement each other.

[C] To be identical with each other.

[D] To be contradictory to each other.

30. The author examines the practice of grade forgiveness by

[A] assessing its feasibility.

[B] analyzing the causes behind it.

[C] comparing different views on it.

[D] listing its long-run effects.

五、标注结构结构说明约定及标注示例

（1）数据结构说明及约定

由于题目数据结构多变不固定，因此计划使用 xml 结构描述题目的内容，并保留原始结构信息。

数据主体结构由两部分构成：

字段名	说明
meta	meta 用于存储题目的所有属性信息，如：来源、语种，文本长度等
content	存储题目的文本内容，包括大小题题干、解析、答案等

考虑到未来对题目进行归集、数据溯源或进行进一步的清洗等操作需要，meta 字段二级展开如下：

字段名	type	必填	说明	备注参考
version	string	是	版本名，不限于版本号，可用于记录该条所在的处理版本或批次	
subject	string	是	门类：**（仅能填写以下门类）*哲学、经济学、法学、教育学、文学、历史学、理学、工学、农业、医学、管理学、艺术类、农林牧渔、资源环境与安全、能源动力与材料、土木建筑、水利、装备制造、生物与化工、轻工纺织、食品药品与粮食、交通运输、电子与信息、医药卫生、财经商贸、旅游、文化艺术、新闻传播、教育与体育、公安与司法、公共管理与服务、农林牧渔、资源环境与安全、能源动力与材料、土木建筑、水利、装备制造、生物与化工、轻工纺织、食品药品与粮食、交通运输、电子与信息、医药卫生、财经商贸、旅游、文化艺术、新闻传播、教育与体育、公安与司法、公共管理与服务。*** 数据标注到**英语四六级试卷或者考研英语试卷、雅思、托福**等外语考试卷子的时候，subject 标签请写成英语；**考研政治试卷** subject 标签请写成政治；**考研数学** subject 标签请写成数学；**计算机**的写为计算机即可。其他科目的试卷按照专业门类表，进行相应分类即可。	大学专业分类表

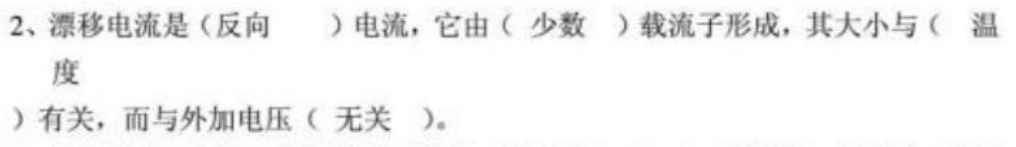
data_type	string	是	原始数据的具体来源，分为教材、教辅、试卷；后续可根据情况拓展，如增加真题、模拟等，区分题目质量	
pub_year	string	否	题目所属试卷的年份，若为教材等无明确年份的可不填；若 pdf 文件名中包含年份，一定要写	
full_name	string	是	书籍或试卷的全名	
chapter	string	否	若为书籍，可写入章节名称	
language	string	是	语料主要语种的标识，如 zh、en 等（繁简体区分按需）	
source	string	是	应该存放该套试卷/教辅/教材的 pdf 网络源路径	
quiz_type	string	是	题型，目前分为 6 类：解答题、选择题、判断题、填空题、做图题、其它题。当一道题目包含多种题型的小题时，该字段写入：复合题	
is_explained	string	是	是否包含答案解析或解题思路分析，填 True 或 False	
is_answered	string	是	是否有答案，填 True 或 False	
pub_time	string	否	发布时间戳，UTC 时间，指数据最原始版本的时间，因为最原始版本后面还可能会有多轮处理	
process_time	string	是	流水线处理时间戳，UTC 时间	
additional_info	xml-string	否	留个活口，方便留备注、额外字段什么的	

content 字段用于存储题目的详细信息，可根据题目结构灵活调整，二级展开如下：

字段名	type	必填	说明
question	string	是	主题干文本内容，若为单题结构，则为问题本身。
quiz	string	否	小题题干，里面可增加参数 type，用于 meta - quiz_type 为复合题时，额外描述题型，题型与 meta - quiz_type 中五大题型一致
id	string	否	题号，标明是在所在大题中的第几小题
exp	string	否	题目解析，若为小题解析，根据小题 id 规则迁移 id
ans	string	否	题目答案，若为小题答案，根据小题 id 规则迁移 id。若答案中同时存在多个答案，用[space]隔开
img	string	是	若题目的题干、解答中包含图标，则需要将每一个图标单独切图，并在该字段放入图片文件路径。多个图片从 1 开始自增，里面可增加参数 src，用于与对应的题干、答案或解析标签进行映射

（2）标注示例

A.单题




```XML
<meta>
    <version>1.0<version/>
    <subject>电子与信息</subject>
    <data_type>试卷</data_type>
    <pub_year>2016</pub_year>
    <full_name>2016 年-天津工业大学-模拟电子技术期末考试模拟题一
</full_name>
```

```
    <language>zh</language>
    <source>xxxxxx</source>
    <quiz_type>填空题</quiz_type>
    <is_explaned>False</is_explaned>
    <is_answered>True</is_answered>
    <pub_time>2023-09-08 19:00:00</pub_time>
    <process_time>2023-09-08 20:00:00</process_time>
    <additional_info>
        <附加字段示例 1></附加字段示例 1>
        <附加字段示例 2></附加字段示例 2>
    </additional_info>
</meta>
<content>
    <question>2、漂移电流是（ ）电流，它由（ ）载流子形成，其大小与（ ）有关，而与外加电压（ ）。</question>
    <ans>反向[space]少数[space]温度[space]无关</ans>
    <img src="question">/path/to/x.png</img>
</content>
```

B.大题小题的标注规则

1. 大题我们用 <question> 来标记。
2. 小题我们用 <quiz> 来标记。
3. 如果没有 quiz 的话， 那么 <question> 需要和 <ans> 一一对应。
4. 如果有 quiz 的话， 那么<question> 同级目录下不需要有 <ans> 。 所有 <ans> 在 <quiz> 下和 <content > 一一对。
5. Quiz 的图片标注规则： 所有 quiz 都需要标注图片 source 来源标注。
6. 一道 quiz 题如果有多个空格。 比如下面，对应的答案部分用 [space]隔开

```XML
内容分析包括（）和 （）哪几个模块？
```

```XML
<ans> A [space] B </ans>
```

7. **大小题拆分逻辑：**
    a. 对于一道大题里的多个小问，排版相对比较密集的 **（一行）**， 不需要拆分

成 quiz 。

b. 对于大题多问（多行），**排版的时候有着留白以回答大量 ans 内容的题目，** 需要拆分成 quiz 。

c. 对于涉及到多页， 如完形填空， 阅读理解等题目， 需要拆分成 quiz 。

d. 如果没有涉及到小题有图片之类的， **quiz 不需要标记 <image> . (quiz.image 是一个 optional)**

e. **具体情况请实际标注过程中和项目经理保持密切沟通**

下面是一些样例:

例 a ： 此道汉语阅读题排版较为密集，不需要拆分成 quiz。

573. 一个句子从不同角度看可以属于多种句型。"松花江被晚霞照得通红"这个句子，可以看成：

①被动句 （ ）

②主谓句 （ ）

③形容词性谓语句 （ ）

④单句 （ ）

⑤动词性谓语句 （ ）

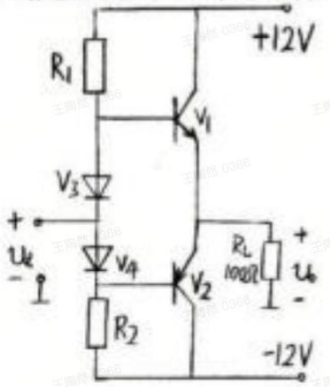
例 b： 此物理应用题有多问，并且排版的时候有着留白以回答大量的 ans 内容，该题目就需要拆分成 quiz。

4、电路如图所示，设 $U_{CES}=0$ 试回答下列问题：（6 分）

（1）ui=0 时，流过 RL 的电流有多大？

（2）若 V3、V4 中有一个接反，会出现什么后果？

（3）为保证输出波形不失真，输入信号 ui 的最大幅度为多少？管耗为多少？




```xml
<?xml version="1.0" encoding="UTF-8"?>
<body>
    <body1>
        <meta>
            <version>1.0</version>
            <subject>电子与信息</subject>
            <data_type>试卷</data_type>
            <pub_year>2016</pub_year>
            <full_name>2016 年-天津工业大学-模拟电子技术期末考试模拟题一</full_name>
            <language>zh</language>
            <source>/path/to/x.png</source>
            <quiz_type>问答题</quiz_type>
            <pub_time>2023-09-14 20:36:51</pub_time>
            <process_time>2023-09-14 22:11:23</process_time>
            <is_explaned>False</is_explaned>
            <is_answered>True</is_answered>
        </meta>
        <content>
            <question>4、电路如图所示,设 U_{CES}  试回答下列问题;(6 分) </question>
            <img src="question">/path/to/x.png</img>
            <img src="question">/path/to/x.png</img>
            <quiz>
            <id>1</id>
```

```
            <content>(1)ui=0\text{ 时,流过 RL 的电流有多
大}</content>
            <img src="question">/path/to/x.png</img>
            <ans> (1) ui=0 时 R_L 电流为零 </ans>
            </quiz>
            <quiz>
            <id>2</id>
            <content>(2)\text{ 若 V3、V4 中有一个接反,会出现
什么后果}?(</content>
            <img src="question">/path/to/x.png</img>
            <ans>2) V3、V4 有一个反接电路不能工作 </ans>
            </quiz>
            <quiz>
            <id>3</id>
            <content>（3）\text{ 为保证输出波形不失真,输入信
号 ui 的最幅度为多少? 管耗为多少}?</content>
            <img src="question">/path/to/x.png</img>
            <ans> 3) ui=0 时 R_L 电流为零 </ans>
            </quiz>
        </content>
    </body1>
</body>
```

例 c： 英语题的所有小题全部按 quiz 格式分别标注出来。

***英语题：question 里面 img 来源需要放上英语阅读原文图片以及英语小题的整张图片如下所示的两张图；而英语题的 quiz 标注则需要对应的每一道小题的单独 img 来源。**

Text 2

Grade inflation – the gradual increase in average GPAs (grade-point averages) over the past few decades – is often considered a product of a consumer era in higher education, in which students are treated like customers to be pleased. But another, related force – a policy often buried deep in course catalogs called "grade forgiveness" – is helping raise GPAs.

Grade forgiveness allows students to retake a course in which they received a low grade, and the most recent grade or the highest grade is the only one that counts in calculating a student's overall GPA.

The use of this little-known practice has accelerated in recent years, as colleges continue to do their utmost to keep students in school (and paying tuition) and improve their graduation rates. When this practice first started decades ago, it was usually limited to freshmen, to give them a second chance to take a class in their first year if they struggled in their transition to college-level courses. But now most colleges, save for many selective campuses, allow all undergraduates, and even graduate students, to get their low grades forgiven.

College officials tend to emphasize that the goal of grade forgiveness is less about the grade itself and more about encouraging students to retake courses critical to their degree program and graduation without incurring a big penalty. "Ultimately," said Jack Miner, Ohio State University's registrar, "we see students achieve more success because they retake a course and do better in subsequent courses or master the content that allows them to graduate on time."

That said, there is a way in which grade forgiveness satisfies colleges' own needs as well. For public institutions, state funds are sometimes tied partly to their success on metrics such as graduation rates and student retention – so better grades can, by boosting figures like those, mean more money. And anything that raises GPAs will likely make students – who, at the end of the day, are paying the bill – feel they've gotten a better value for their tuition dollars, which is another big concern for colleges.

Indeed, grade forgiveness is just another way that universities are responding to consumers' expectations for higher education. Since students and parents expect a college degree to lead to a job, it is in the best interest of a school to turn out graduates who are as qualified as possible – or at least appear to be. On this, students' and colleges' incentives seem to be aligned.

微信公众号-世纪高教在线-回复关键词-黄皮书领课-即可免费学习考研英语各题型解题技巧

26. What is commonly regarded as the cause of grade inflation?

[A] The change of course catalogs.

[B] Students' indifference to GPAs.

[C] Colleges' neglect of GPAs.

[D] The influence of consumer culture.

XML

```
 <?xml version="1.0" encoding="UTF-8"?>
<body>
    <body1>
        <meta>
            <version>1.0</version>
            <subject>英语</subject>
            <data_type>试卷</data_type>
            <pub_year>2016</pub_year>
            <full_name>2019 年考研英语一试卷</full_name>
            <language>en</language>
            <source>http://159.75.220.17:8080/dev-
api/tools/view?fileId=1701888659688534016</source>
            <quiz_type>选择题</quiz_type>
            <pub_time>2023-09-14 20:44:33</pub_time>
            <process_time>2023-09-14 22:11:19</process_time>
            <is_explaned>False</is_explaned>
            <is_answered>False</is_answered>
        </meta>
        <content>
            <question>Text 2 Grade inflation – the gradual
increase in average GPAs (grade-point averages) over the past few
decades – is often considered a product of a consumer era in
higher education, in which students are treated like customers to
be pleased. But another, related force – a policy often
buried deep in course catalogs called “grade forgiveness”
– is helping raise GPAs.
Grade forgiveness allows students to retake a course in which they
received a low grade, and the most recent grade or the highest
grade is the only one that counts in calculating a student’s
overall GPA.
The use of this little-known practice has accelerated in recent
years, as colleges continue to do their utmost to keep
students in school (and paying tuition) and improve their
graduation rates. When this practice first started decades ago, it
was usually limited to freshmen, to give them a second chance to
take a class in their first year if they struggled in their
transition to college-level courses. But now most colleges,
save for many selective campuses, allow all undergraduates,
and even graduate students, to get their low grades forgiven.
College officials tend to emphasize that the goal of grade
forgiveness is less about the grade itself and more about
encouraging students to retake courses critical to their degree
program and graduation without incurring a big penalty.
```

“Ultimately,” said Jack Miner, Ohio State University’s registrar, “we see students achieve more success because they retake a course and do better in subsequent courses or master the content that allows them to graduate on time.”
That said, there is a way in which grade forgiveness satisfies colleges’ own needs as well. For public institutions, state funds are sometimes tied partly to their success on metrics such as graduation rates and student retention – so better grades can, by boosting figures like those, mean more money. And anything that raises GPAs will likely make students – who, at the end of the day, are paying the bill – feel they’ve gotten a better value for their tuition dollars, which is another big concern for colleges.
Indeed, grade forgiveness is just another way that universities are responding to consumers’ expectations for higher education. Since students and parents expect a college degree to lead to a job, it is in the best interest of a school to turn out graduates who are as qualified as possible – or at least appear to be. On this, students’ and colleges’ incentives seem to be aligned.
.</question>
    <img src="question">/path/to/image</img>
    <quiz>
    <id>1</id>
    <content>26. What is commonly regarded as the cause of grade inflation?
[A] The change of course catalogs.
[B] Students’ indifference to GPAs.
[C] Colleges’ neglect of GPAs.
[D] The influence of consumer culture.</content>
    <img src="quiz" > /path/to/image </image>
    <ans>A</ans>
    </quiz>
    <quiz>
    <id>2</id>
    <content>27. What was the original purpose of grade forgiveness?
[A] To help freshmen adapt to college learning.
[B] To maintain colleges’ graduation rates.
[C] To prepare graduates for a challenging future.
[D] To increase universities’ income from tuition.</content>
    <img src="quiz" > /path/to/image </image>
    <ans>A</ans>
    </quiz>

```
    <quiz>
    <id>3</id>
    <content>28. According to Paragraph 5, grade forgiveness
enables colleges to
[A] obtain more financial support.
[B] boost their student enrollments.
[C] improve their teaching quality.
[D] meet local governments' needs.</content>
    <img src="quiz" > /path/to/image </image>
    <ans>B</ans>
    </quiz>
    <quiz>
    <id>4</id>
    <content>
29. What does the phrase "to be aligned" (Line 5, Para. 6) most
probably mean?
[A] To counterbalance each other.
[B] To complement each other.
[C] To be identical with each other.
[D] To be contradictory to each other.</content>
    <img src="quiz" > /path/to/image </image>  # optional
    </quiz>
    <ans>
    <id>4</id>
    <content>A</content>
    </ans>
    </content>
    </body1>
</body>
```

C. Latex 的标注规则

1. 涉及到特殊的识别符号 比如 \div , power , 行列式， limit 等用 latex 来标注。
2. 多行数学公式，用 latex 。
3. Latex 标签 <latex> </latex> 来做标签， <latex>x_2</latex>
4. 单个阿拉伯数字 比如 1， 2，3，a>0，a=1 等等用 utf-8 即可。 像 a=1 这种，简单的就直接敲；y=(x-1)也可以使用 UTF-8 打出来，y=(x-1)(x-2)，也可以用 UTF-8 进行转写。
5. &gt &lt 请用 > < 替代

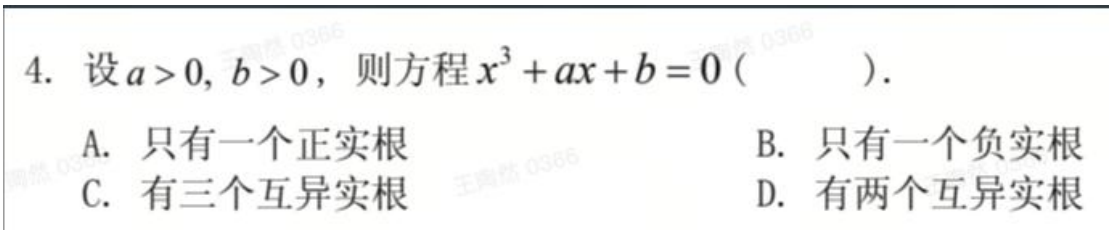
例如：以下例子全部都是正确的

<latex>a^3+b^2=1</latex>
<latex>a^3+b^2=1</latex>(a>0)
<latex>a^3+b^2=1</latex>，<latex>a^4+c^3=0</latex>

公式查询网址：https://www.latexlive.com/

简单可之前输入的公式无需转写，如 a>o,复杂且无法直接输入的公式需要转写

latex 标注样例，如下所示：


4. 设 $a>0$, $b>0$，则方程 $x^3+ax+b=0$（　　）.

A. 只有一个正实根　　B. 只有一个负实根
C. 有三个互异实根　　D. 有两个互异实根

```XML
<?xml version="1.0" encoding="UTF-8"?>
<body>
    <body1>
        <meta>
            <version>1.0</version>
            <subject>数学 A</subject>
            <data_type>试卷</data_type>
            <pub_year>2016</pub_year>
            <full_name>高数 A</full_name>
            <language>zh</language>
            <source>/path/to/image</source>
            <quiz_type>选择题</quiz_type>
            <pub_time>2023-09-14 20:47:28</pub_time>
            <process_time>2023-09-14 21:32:45</process_time>
            <is_explaned>False</is_explaned>
            <is_answered>True</is_answered>
        </meta>
        <content>
            <question>4. 设 a > 0,  b > 0 , 则方程
<latex>x^3+ax+b=0</latex>(      ).
A. 只有一个正实根                    B. 只有一个负实根
C. 有三个互异实根                    D. 有两个互异实根
            </question>
            <img src="question">/path/to/image</img>
            <ans>4. B</ans>
        </content>
    </body1>
```

```
</body>
```

D.answer 的标注规范

1. 如果题目没有答案的话，需要按以下的标注方法标注。有 quiz 的情况下不需要标注.

```
XML
<ans></ans>
```

2. 如果没有 quiz 的情况，<question> 需要和 <ans> 一一对应，ans 无 id；如果有 quiz 。<quiz> 需要 和<ans> 一一对应，且 id 必须一致。

```
XML
 <?xml version="1.0" encoding="UTF-8"?>
<body>
    <body1>
        <meta>
            <version>1.0</version>
            <subject>英语</subject>
            <data_type>试卷</data_type>
            <pub_year>2016</pub_year>
            <full_name>2019 年考研英语一试卷真题</full_name>
            <language>zh</language>
            <source>/path/to/image</source>
            <quiz_type>问答题</quiz_type>
            <pub_time>2023-09-14 20:44:33</pub_time>
            <process_time>2023-09-14 22:11:19</process_time>
            <is_explaned>False</is_explaned>
            <is_answered>False</is_answered>
        </meta>
        <content>
            <question>xx </question>
    <img src="question">/path/to/image</img>
    <quiz>
    <id>1</id>
    <content>xx </content>
    <img src="quiz" > /path/to/image </image>
    <ans>
    <id>1</id>
    <content>A</content>
    </ans>
```

```
    </quiz>
    <quiz>
    <id>2</id>
    <content> xx </content>
    <img src="quiz" > /path/to/image </image>
    <ans>
    <id>2</id>
    <content>A</content>
    </ans>
    </quiz>
    <quiz>
    <id>3</id>
    <content> xx </content>
    <img src="quiz" > /path/to/image </image>
    <ans>
    <id>3</id>
    <content>A</content>
    </ans>
    </quiz>
    <quiz>
    <id>4</id>
    <content> xx </content>
    <img src="quiz" > /path/to/image </image>  # optional
    <ans>
    <id>4</id>
    <content>A</content>
    </ans>
    </quiz>
    </content>
    </body1>
</body>
```

E.多个答案的转写格式规范

多个答案用 [space] 隔开。 例如有三个答案，“A[space]B[space]C[space]”

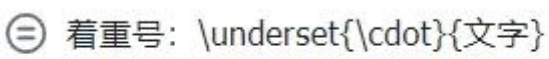
```
Bash
<?xml version="1.0" encoding="UTF-8"?>
<body>
    <body1>
        <meta>
            <version>1.0</version>
```

```
            <subject>文学</subject>
            <data_type>教辅</data_type>
            <pub_year>2016</pub_year>
            <full_name>现代汉语千题解</full_name>
            <language>zh</language>
            <source>http://159.75.220.17:8080/dev-
api/tools/view?fileId=17018886701659054O8</source>
            <quiz_type>判断题</quiz_type>
            <pub_time>2023-09-14 20:25:58</pub_time>
            <process_time>2023-09-14 21:33:38</process_time>
            <is_explaned>False</is_explaned>
            <is_answered>True</is_answered>
        </meta>
        <content>
            <question>573.一个句子从不同角度看可以属于多种句型。"松花
江被晚霞照得通红"这个句子，可以看成：
①被动句（）
②主谓句（）
③形容词性谓语句（）
④单句（）
⑤动词性谓语句（）</question>
            <img src="question">http://159.75.220.17:8080/dev-
api/tools/view?fileId=1702298767744569344</img>
            <ans>√ [space]√ [space]× [space]√ [space]√</ans>
        </content>
    </body1>
</body>


#
```

F.关于题目中的着重号

正常转写即可，题什么样，转写下来就是什么样，要是 UTF8 表示不出来，就用 latex 公式。



例如：

```
\underset{\cdot}{好}
```

输出区域 Output

G.大于号小于号与尖括号的转写问题

大于号小于号与尖括号表示的方法是一致的，所以我们不再区分大于号、小于号和尖括号；不管是大于小于号还是尖括号都用&gt、&lt 进行转写

H.ans 的配图标注规范

ans 如果需要有配图才能回答问题的，直接在 ans 的下一行加一层 img 即可，不需要包裹进 ans 字段里。

```Bash
<body>
  <body1>
    <meta>
      <language>en</language>
      <full_name>612 基础英语 2015_2</full_name>
      <pub_year>2015</pub_year>
      <data_type>试卷</data_type>
      <subject>英语</subject>
      <version>1.0</version>
      <source>http://43.136.104.160:8080/dev-
api/tools/view?fileId=1705508874053775360</source>
      <quiz_type>选择题</quiz_type>
```

```
      <pub_time>2015</pub_time>
      <process_time>2023-10-12 17:10:37</process_time>
      <is_answered>False</is_answered>
      <is_explaned>False</is_explaned>
    </meta>
    <content>
      <question>3. Which of the following italicized parts is used asan object complement? A. The front door remained locked. B. The boy looked disappointed. C. Nancy appeared worried. D. He seemed to have no money left.</question>
      <img src="question">0.png</img>
      <ans>4.A</ans>
      <img src="ans">4.png</img>
    </content>
  </body1>
</body>
```

六、标注数据交付

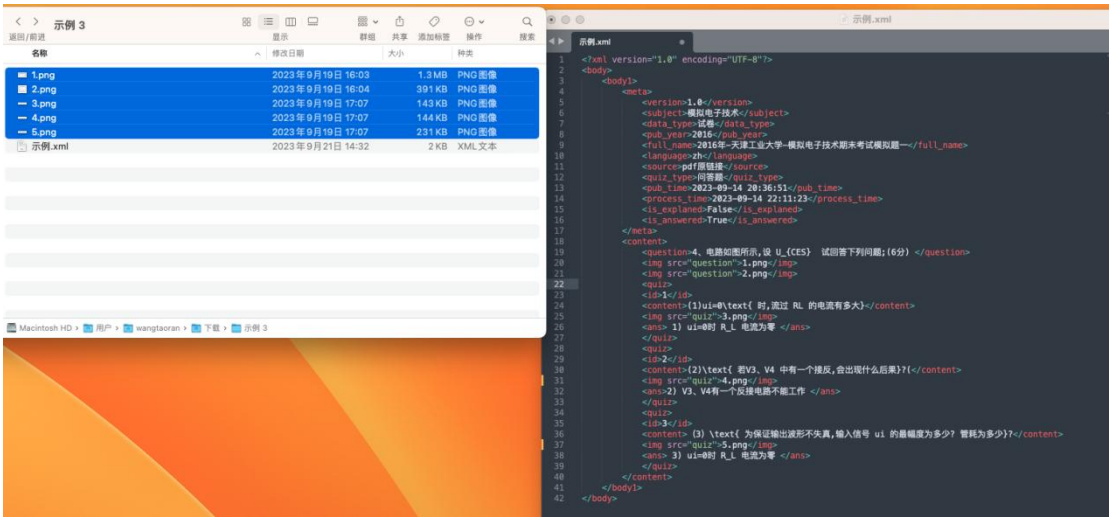
参考文档：大学拍照搜题数据交付格式

1、交付方式：

交付时，应当交付 xml 数据及图片数据，上传至相应的 oss 路径中。

每份试卷或教辅数为一个文件夹，其中包含 img 目录及平铺的 xml 文件，每个 xml 文件中放一道题目的信息。xml 中的图片路径应当为相对路径。具体可参见以下示例。

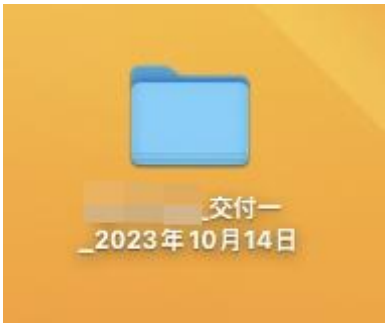
[示例 3.zip]

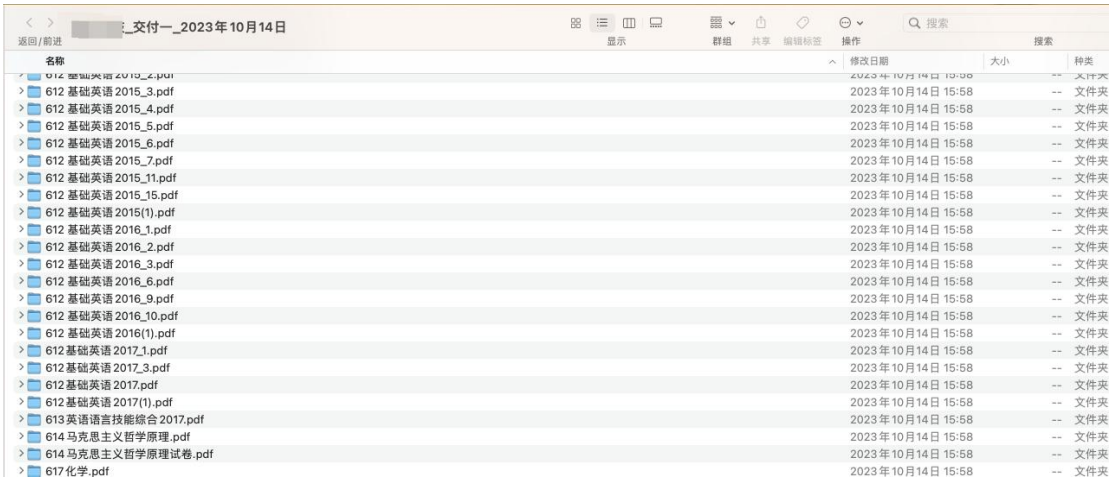


2、交付格式：

1、填写自己家飞书交付数据表格，此表格已发送至各家项目经理；

2、交付文件夹的形式为：交付的 zip 总压缩包（XXX_交付_年月日）——pdf 文件夹——PDF 中每道题的文件夹（问题在试卷中的题号_时间戳）——图片和 XML





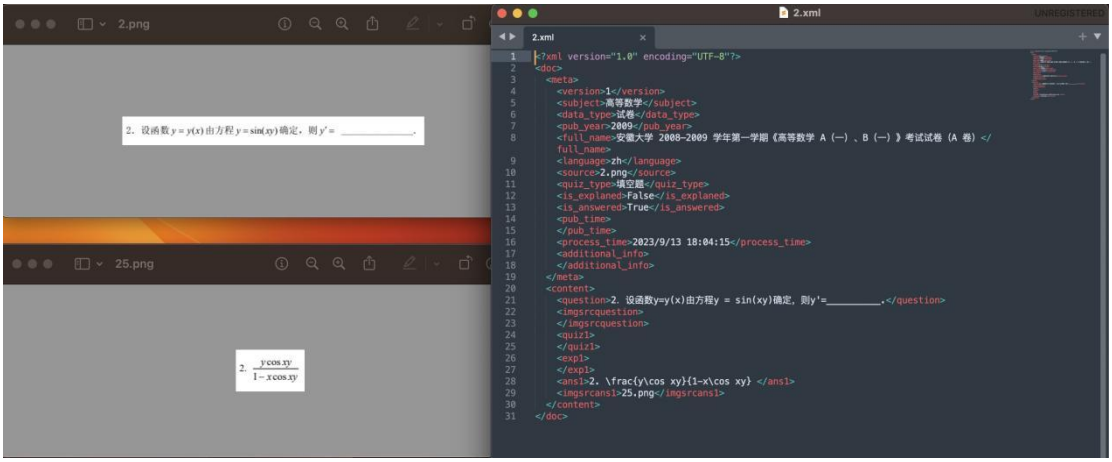
3、根据飞书上的交付表格，在原提起需求邮件的基础上正式回复数据交付邮件（包括所标数据总量，及题型分类）。

七、常见错误案例展示

以下均为错误举例，不要引用！！！！！

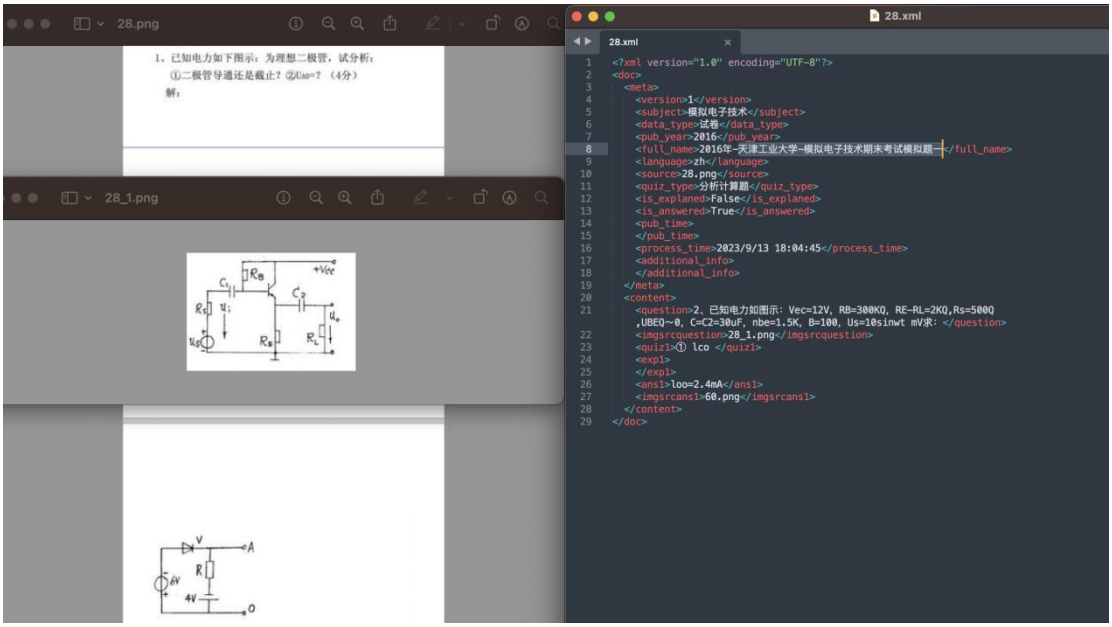
1、多切图

数学题中将答案也切了图，标注进了字符串里。



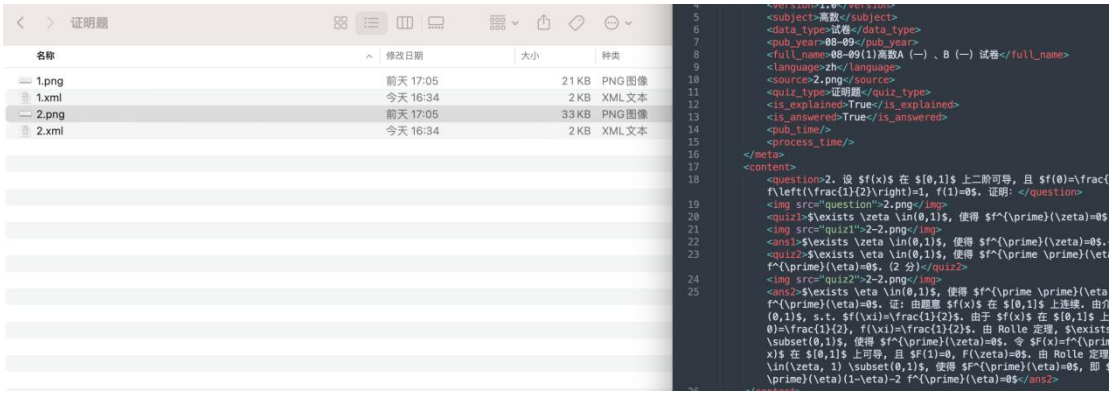
2、错切图

切图错误



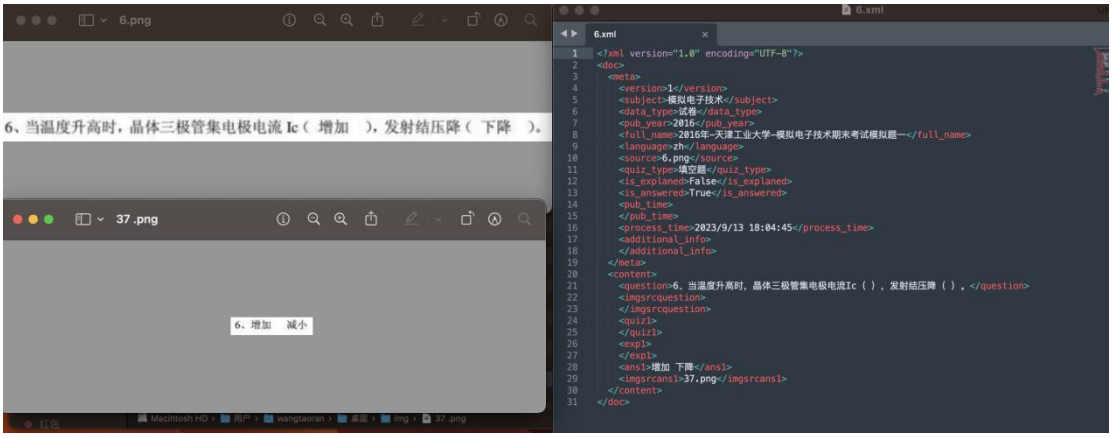
3、少切图

试卷证明题 2 缺少 2-2 的图片



4、缺少图片的 source 字段

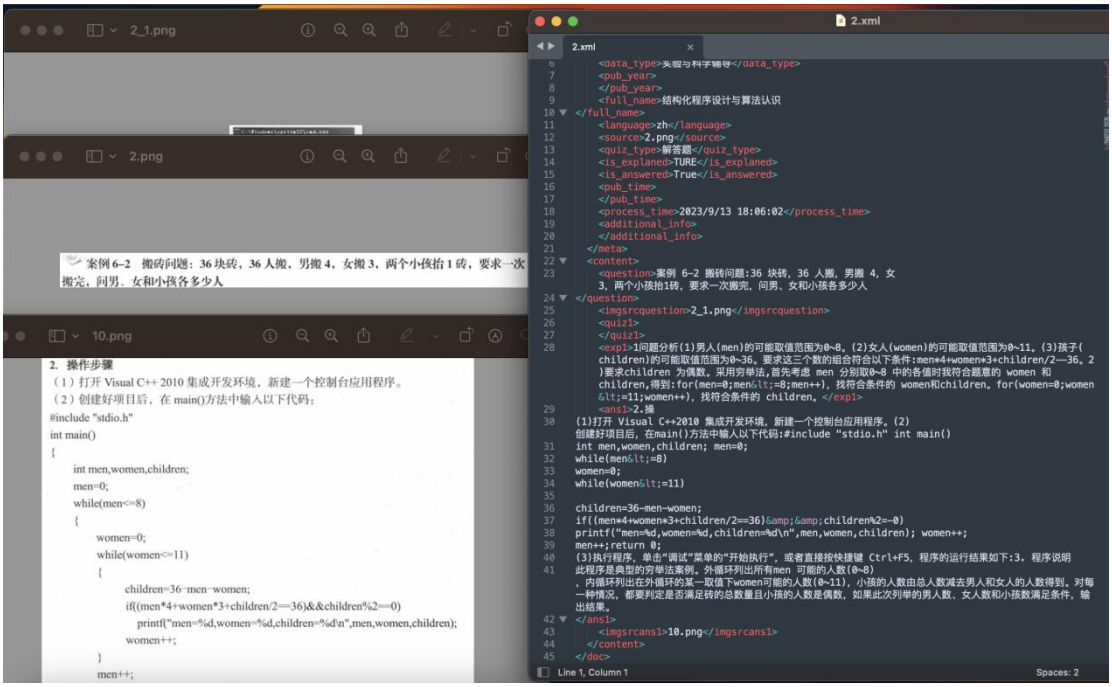
欠缺 question 的 source 字段。



5、缺少 latex 转写

```xml
<?xml version="1.0" encoding="UTF-8"?>
<doc>
  <meta>
    <version>1</version>
    <subject>高等数学</subject>
    <data_type>试卷</data_type>
    <pub_year>2009</pub_year>
    <full_name>安徽大学 2008—2009 学年第一学期《高等数学 A（一）、B（一）》考试试卷（A 卷）</full_name>
    <language>zh</language>
    <source>20.png</source>
    <quiz_type>综合分析题</quiz_type>
    <is_explaned>False</is_explaned>
    <is_answered>True</is_answered>
    <pub_time>
    </pub_time>
    <process_time>2023/9/13 18:04:15</process_time>
    <additional_info>
    </additional_info>
  </meta>
  <content>
    <question>2.设曲线y=x^{3}+3ax^{2}+3bx+c在x=-1处取得极大值,点(0,3)是拐点,求 a,b,c.</question>
    <imgsrcquestion>
    </imgsrcquestion>
    <quiz1>
    </quiz1>
    <exp1>
    </exp1>
    <ans1>解: 函数 y=x^{3}+3 a x^{2}+3 b x+c 在定义域内可二阶可导. 由于曲线在 x=-1 处取到极值, 由 Fermat
    定理, y^{\prime}(-1)=0 . 即 3-6 a+3 b=0 由于 (0,3) 是曲线的拐点, 则 y^{\prime \prime}(0)=0, y(
    0)=3 , 即 a=0, c=3 , 带入得到 b=-1 .</ans1>
    <imgsrcans1>43.png</imgsrcans1>
  </content>
</doc>
```

6、题目文字转写有误



7、转写字段错误

```
        <process_time>2023-09-14 21:32:35</process_time>
        <is_explaned>False</is_explaned>
        <is_answered>True</is_answered>
      </meta>
      <content>
        <question>2. 设 f (x)在[0,1]上二阶可导, 且 f (0)=\frac{1}{2}, f(\frac{1}{2} )=1,f(1)=0.
证明: </question>

        <quiz>(1) 3ζ ∈(0,1), 使得f'(ζ)=0. (4 分) </quiz>

        <img src="question">http://159.75.220.17:8080/dev-api/tools/view?fileId=1702305375815995392</img>
        <ans>\begin{aligned}&amp;\text{证: 由题意 }f(\mathrm{x})\text{ 在}[0,1]\text{ 上连续,}\quad\text{由介值定理,}\quad\exists\xi\in(0,1),\mathrm{~s.t.~}f
(\xi)=
<latex>\frac12.</latex>
\\&amp;\text{由于 }f(\mathrm{x})\text{ 在}[0,1]\text{ 上可导,}f(0)=
<latex>\frac12.</latex>
,\:f(\xi)=
<latex>\frac12.</latex>
.\quad\text{由 Ro11e 定理,}\:\exists\zeta\in(0,\xi)\subset(0,1)\:,\&amp;\text{使得 }f^{\prime}(\xi)=0.\&amp;\quad</ans>

        <quiz>(2) 3η ∈(0,1), 使得(f'')(1-η)-2f'(η)=0. (2分) </quiz>

        <ans>\text{令}F(\mathrm{x})=f^{\prime}(\mathrm{x})(\mathrm
<latex>{l-x})^2</latex>
,\:\text{ 显然 }F(\mathrm{x})\:\text{在}[0,1]\:\text{上可导},\quad\text{且}\:F(1)=0,\:F(\zeta)=0\:.amp;\text{ 由 Ro11e 定理,}\quad\exists\eta\in(\zeta,1)\subset(0,1),
\text{使得 }F^{\prime}(\eta)=0\:,\:\text{ 即 }
</latex>
\:f^{\prime\prime}(\eta)(1-\eta)-2\:f^{\prime}(\eta)=0
</latex>
\end{aligned}</ans>
        <img src="ans">http://159.75.220.17:8080/dev-api/tools/view?fileId=1702308559032684544</img>
      </content>
    </body1>
</body>
```

8、题目和答案判断错误

题干与答案（如有）应该分开标注，题是题，答案是答案，题里面不能包含答案。

9、图片来源字符格式有误

```
        <question>Part A 51. Directions: Suppose you are working for the &amp;quot;Aiding Ru
        <img src="question">.png</img>
        <ans>Dear Sir ,  It is appreciated to know from your  letter of that you applied
    </content>
```

八、补充说明

1) 文件夹命名格式变更：

正式交付文件夹命名格式如下：

pdf 文件夹——PDF 中每道题的文件夹（问题在试卷中的题号_时间戳）——图片和 XML

例如：pdf 文件夹命名 ---大题号-小题号-流水线处理时间戳（可以是时间+任意数字）----图片和 xml

注意时间戳一定是唯一的，确保每个文件夹命名都是独一无二的

例如：大题号（数字）+小题号（数字）+流水线时间戳年月日（必须和程序内一致）+任意数字（5-8 位） 1-1-20231023175911122233

2）关于 img 引用为 question 和 quiz 区别：

若是英语的完型填空和阅读理解，为 quiz

若是其他题进行拆分则为 question

3） 1.关于答案中的题号问题： 场景一：题目是一、XXX 答案是：1） 格式保持一致 场景二：题目是一、XXXX 答案没题号，直接是答案内容：不需要写题号

2.补充知识点： ans 如果需要有配图才能回答问题的，直接在 ans 的下一行加一层 img 即可，不需要包裹进 ans 字段里。

4）名词解释可以写为解答题

5）表格中没有的课程根据题目判断学科，可以进行百度，并选择最符合的门类

6）无意义的空格可不打，英文单词中的空格必须打

7）公式转写中 latex 位置的判断标准，仅需要将转写的部分用 latex 标签包裹，判断的标准为在 latex 工具中输出区域的呈现方式与 pdf 文件保持一致