

# “大雾实验工具”的开发

## 《程序设计进阶与实践》大作业报告

姓名（组长）	孙旭磊	学号	PB21000270
姓名（组员1）	张学涵	学号	PB21000079
姓名（组员2）	秦沁	学号	PB21111630
姓名（组员3）	赵弈	学号	PB21000033
姓名（组员4）	鲍政廷	学号	PB21111741
项目名称	大雾实验工具——绘制图像&计算不确定度&生成计算公式		

## 1 项目需求分析

大物实验在[评课社区](#)、[知乎](#)等网站上一直饱受争议，有多名同学指出实验报告撰写耗时长、专业作图软件难以使用、Word 中打数学公式麻烦等问题。鉴于此，曾经有学长在开发过一款[大物实验数据处理工具](#)，这是非常好的创意。但是，这款软件入门成本太高，故了解它的人很少。

本小组开发的[大雾实验工具](#)是一款网页应用，无需安装任何软件，更不需要有编程基础，没有任何学习成本。本工具的目标用户是中国科学技术大学大一本本科生，着力于解决其撰写实验报告时最耗时的三件事情，即“绘制图像”“计算不确定度”“在电脑上书写公式”。

当然，一些高级软件也能出色地完成上述的本工具的功能，如专业绘图软件 Origin，专业计算软件 Matlab 等。但我们的项目不是去取代这些强大的软件，而是将它们本地化。这些软件功能繁多，故学习成本相对较高，但我们的软件为每一个大物实验都写了专门的处理工具，封装到只需要用户上传数据表格的程度。相比动辄几个 GB 的专业软件来说，我们的工具更加友好，更加便捷，更加有针对性——更加有效。

## 2 项目功能设计

### 2.1 总体功能说明

大雾实验工具是本组成员 2022 春季学期程序设计进阶与实践的大作业项目。本工具通过腾讯云服务器搭建于网页平台，支持任何设备自由访问。传入实验数据后，本工具立刻完成绘制图像、计算不确定度、生成计算公式等一系列操作，并将最终结果整理成一份 Word 文档，下载后即可直接使用。本实验工具支持一级大物的 25 个实验，如图 1 所示，这大大提升了学生们撰写实验报告的效率。由于本工具只是将传入的实验数据进行自动分析，故其不会造成抄袭、造假等学术不端问题。

### 2.2 具体功能点说明

使用本工具时，用户只需输入他们做实验时测量到的原始数据，而无需任何额外的计算处理，用户所要做的只有按照规定的格式上传 Excel 文档。本工具支持 `xlsx`, `csv` 等各种格式的数据表格。具体而言，每个实验都会有一张示例数据表供用户参考，如图 2 的界面所示。用户也可以直接下载示例数据，并直接在它的基础上进行修改。因此，本工具没有任何学习成本，是一款即点即用、免安装的简单轻应用。

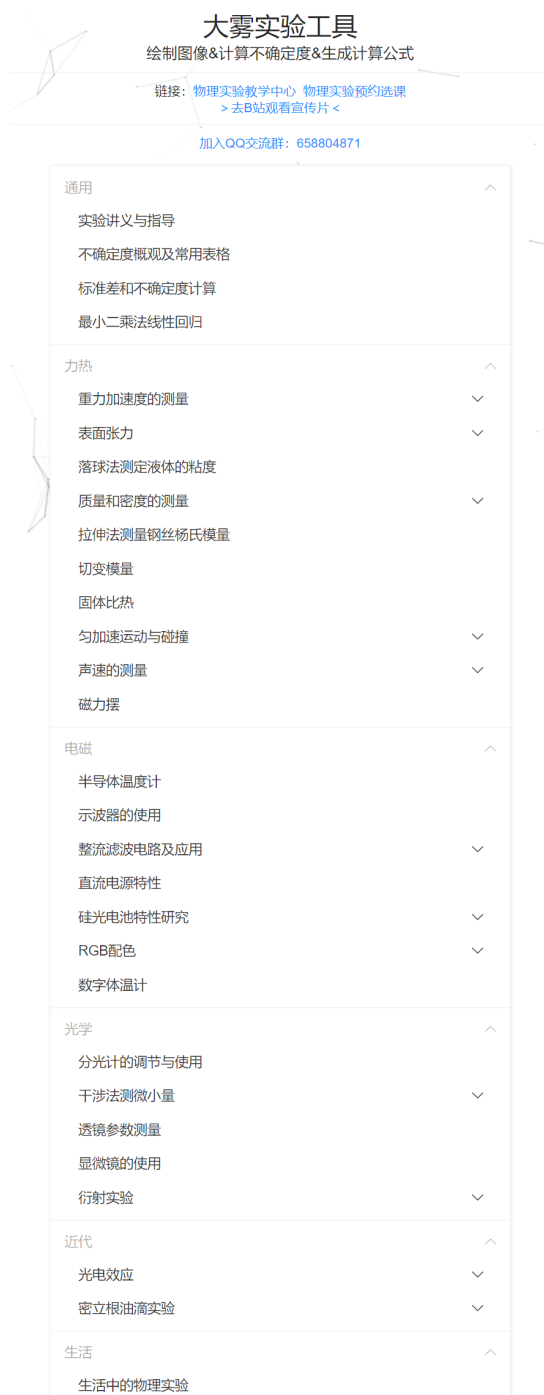


图 1: 大雾实验工具网站主界面

另外，本工具贴心地提供了不确定度表格与通用的计算工具，并且每个实验都附有可在线浏览的实验指导。

## 绘制图像

本工具根据输入的数据以及实验原理，自动生成美观的实验图像，支持平滑去噪、数据拟合、双 y 图等多种图像生成需求，如图 3 所示。

## 计算不确定度与生成计算公式

本工具在生成的 Word 文档中渲染了各种公式，如图 4 所示。用户可以直观看到不确定度每一步的计算过程，并在自己的报告中直接使用这些算式与结果。

在 Word 文档中除了有已经渲染好的公式外，我们还提供了它们的  $\text{\LaTeX}$  源码。这极大方便了用  $\text{\LaTeX}$ , Markdown 等排版实验报告的用户，他们无需手动敲入每一个算式了。

## 2.3 功能点设计细节

本工具后端使用 Python 编写，使用的包与模块如表 1 所示。前端由 HTML 编写，并使用了 Flask Web 应用框架。

### 2.3.1 图像绘制

图像由 Matplotlib 绘制。我们的规范如下：

- 面向绘图对象作图: `fig, ax = matplotlib.pyplot.subplots()`
- 设置副刻度为主刻度的一半，主刻度为默认: `ax.xaxis.set_minor_locator(matplotlib.ticker.AutoMinorLocator(2))`



图 2: “拉伸法测钢丝杨氏模量”的工具界面

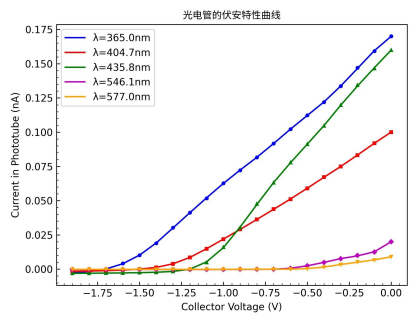


图 3: 平滑连接的光电效应伏安特性曲线

钢丝直径  $d$  的平均值:

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i = \frac{0.291 + 0.292 + 0.293 + 0.294 + 0.295}{5} \text{ mm} = 0.293 \text{ mm}$$

钢丝直径  $d$  的标准差:

$$s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2} = \sqrt{\frac{(0.291 - 0.293)^2 + (0.292 - 0.293)^2 + (0.293 - 0.293)^2 + (0.294 - 0.293)^2 + (0.295 - 0.293)^2}{5-1}} \text{ mm}$$

$$= 0.0015811 \text{ mm}$$

钢丝直径  $d$  的 B 类不确定度:

$$u_{B,d} = \sqrt{u_{B,d}^2 + u_{B,d}^2} = \sqrt{0.004^2 + 0.005^2} \text{ mm} = 0.0064031 \text{ mm}$$

钢丝直径  $d$  的展伸不确定度:

$$U_{d,P} = \sqrt{\left(1.96 \times \frac{s_d}{\sqrt{n}}\right)^2 + \left(k_P \times \frac{u_{B,d}}{C}\right)^2} = \sqrt{\left(2.78 \times \frac{0.0015811}{\sqrt{5}}\right)^2 + \left(1.96 \times \frac{0.0064031}{3}\right)^2} \text{ mm}$$

$$= 4.6222 \times 10^{-3} \text{ mm}, P = 0.95$$

图 4: 不确定度计算的详细过程

表 1: 本工具使用的全部 Python 包与模块

Python 包或模块	用途
chardet	检测用户上传的数据表格的编码
collections	通过 <code>namedtuple</code> 使代码更清晰
Flask	Web 应用框架
latex2mathml	L <sup>A</sup> T <sub>E</sub> X 代码转换为 MathML 代码
lxml	MathML 代码转 Office MathML
math, numpy	不确定度数字运算
Matplotlib	绘制物理图像
os, random, shutil	后台文件操作与管理
pandas	数据表格处理
python-docx	生成 Word 文档
SciPy	数据拟合
SymPy	不确定度符号运算
time, threading	定时删除生成的 Word 文档
traceback	打印运行错误以便调试

- 刻度朝内: `matplotlib.rcParams["xtick.direction"] = matplotlib.rcParams["ytick.direction"] = "in"`
- 若一张图有且只有一组点线, 则点使用红色 (`color="r"`), 线使用蓝色 (`color="b"`), 且线覆盖在点的上面; 若一张图有多组点线, 则同一组点线的颜色应当相同, 并依次使用蓝(b)、红(r)、绿(g)、紫(m)、橙(orange)、青(c)。
- 点的类型使用实心圆 ("o"), 若一张图有多组点线, 则依次使用实心圆(o)、正方形(s)、上三角(^)、菱形(D)、下三角(v)、星号(\*)。
- 线条粗细使用 `linewidth=1.5`, 点的大小使用 `markersize=3`, 可视数据量、数据组数适当调整, 但应保持统一性。

- 绘制双 y 轴图使用 `matplotlib.axes.Axes` 对象的 `twinx()` 方法。
- 只有一组点线的图，一般不显示图例。
- 图像字体: `SourceHanSansSC-Regular.otf`
- 轴标签和标题中的物理量名称与单位应使用  $\text{\LaTeX}$ 。

### 2.3.2 数据处理

无论使绘制图像时的线性拟合，还是计算不确定度的大小，都绕不开数据处理。我们利用 `pandas`, `SciPy`, `SymPy` 等包自主编写了 `calc.py` 应用程序接口，它提供以下函数：

**科学计数法输出** `numlatex: (num: float,`

`prec: int = 5) -> str`

返回一个数的科学计数法形式的  $\text{\LaTeX}$  代码

`num`: 要转成科学计数法的数字

`prec`: 有效数字位数（默认值：5）

**不确定度计算** `analyse: (data: pandas.DataFrame,`

`delta_b1: float = 0., delta_b2: float`

`= 0., symbol: str = "x", unit: str = "",`

`confidence_C: float = 3., confidence_P:`

`float = 0.95) -> AnalyseData`

计算一组数据的平均值、标准差、不确定度

`data`: 要处理的一组实验数据

`delta_b1`: 仪器最大允差  $\Delta_{\text{仪}}$ （默认值：0）

`delta_b2`: 估读最大允差  $\Delta_{\text{估}}$ （默认值：0）

`symbol`: 数据的物理符号（默认值："x"）

`unit`: 数据的单位（默认值："")

`confidence_C`: 置信系数  $C$ （默认值：3）

`confidence_P`: 置信概率  $P$ （默认值：0.95）

`AnalyseData`: 数据计算结果的集合

**最小二乘法线性回归** `analyse_lsm: (data_X:`

`pandas.DataFrame, data_Y: pandas`

`.DataFrame, symbol_X: str = "X",`

`symbol_Y: str = "Y", unit_m: str = "",`

`unit_b: str = "") -> AnalyseLsmData`

将一组数据用最小二乘法拟合成一条直线

`data_X`: x轴数据（自变量数据）

`data_Y`: y轴数据（因变量数据）

`symbol_X`: 自变量物理符号（默认值："X"）

`symbol_Y`: 因变量物理符号（默认值："Y"）

`unit_m`: 斜率的单位（默认值："")

`unit_b`: 截距的单位（默认值："")

`AnalyseLsmData`: 直线拟合结果的集合

**不确定度合成** `analyse_com: (exp: str, varr:`

`tuple = (), constt: tuple = (), unit:`

`str = "", confidence_P: float = 0.95)`

`-> AnalyseComData`

根据表达式计算物理量的值和不确定度

`exp`: 物理量计算表达式（字符串），为一个物理量=一些物理量（或常量）之积与之商的形式，如  $E=4\pi^2 l/T^2$  代表  $E = \frac{4\pi^2 l}{T^2}$

`varr`: 物理量（元组），元组的每个元素均为元组，该子元组的第1个元素为物理量名，第2个元素为物理量值，第3个元素为其不确定度（默认值：()）

`constt`: 常量（元组），元组的每个元素均为元组，该子元组的第1个元素为常量名，第2个元素为常量值（默认值：()）

`unit`: 要计算的物理量的单位（默认值："")

`confidence_P`: 置信概率  $P$ （默认值：0.95）

`AnalyseComData`: 不确定度合成结果的集合

### 2.3.3 文档生成

Word 文档由 `python-docx` 生成。我们的规范如下：

- 字体使用微软雅黑：`document.styles['Normal'].font.name = "微软雅黑"`
- 文档第一行是实验名称：  
`document.add_paragraph(name())`  
随后注明：  
“【Latex 代码在下面，请向下翻阅】”
- 内容跨度较大的段落之间应当用一个空行。
- 文档中插入的数据一般保留 4 或 5 位有效数字：`"%.5g"%x`，线性拟合的相关系数  $r$  保留 8 位有效数字。
- 若某张图片正好在第 2 页开头，而第 1 页尾部有很多空白区域，为避免误解，应在第 1 页的最后一个段落之后注明“【本文档不只有一页，请向下翻阅】”。
- 插入表格使用 `docx.document.Document` 对象的 `add_table()` 方法。

鉴于不确定度的计算方法是固定的、算法化的，我们利用 `lxml` 等包自主编写了 `insert.py` 应用程序接口，这样只需调用几个函数，就可以在 Word 文档中完成数学算式的渲染与添加。具体可见公式插入 API 的说明文档，这里不再赘述。

## 3 测试、运行情况

### 3.1 测试情况

本程序的每一个实验模块由组员完成后，组长会进行代码审核与测试，如果发现问题则要求

继续修改，直到所有问题被解决后该实验模块才会发布。我们还建立了用户 QQ 群，并即时反馈用户提出的任何问题。

另一方面，各种 API 的编写与模块化编程也让我们的程序在编写过程中更不容易出错，同时规范、统一的码风也让调试变得轻松。

### 3.2 运行情况

本项目从 2022 年 4 月 27 日上线以来，有超过 9000 次的浏览量，总访客数达到了 2100。

最后一次大物实验于 2022 年 6 月 3 日结束。在大物实验结束之前，平均每天约有 50 名同学访问了我们网站，如图 5 所示。考虑到每天只有不到 400 名大一学生做大物实验，本程序的使用率相对较高。用户的平均访问时长为 6 分 36 秒，这说明我们的大雾实验工具非常简明易用。

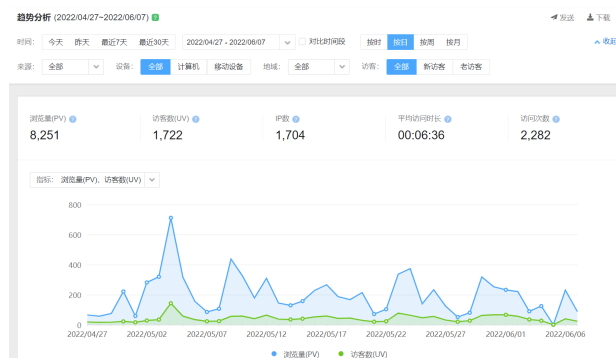


图 5: 4 月 27 日至 6 月 6 日的统计数据

## 4 设计、开发过程中的难点

鉴于我组成员都有一定的编程基础，整个项目中我们并没有遇到较大的困难，主要只有两个方面的小障碍：

第一，对实现的接口不够熟悉。由于我们编写的 API 参数较多，故未参与编写 API 的组员在一开始时需要花一点时间熟悉其调用方式。不过在后期大家也逐渐熟练地掌握了所有 API 的使用方式，而不用再去翻阅帮助文档。

第二，对物理实验本身不够了解。有的物理实验原理复杂、参考资料有限，且实验课程标准在不断变化，这会使程序生成的报告存在瑕疵，影响本项目的声誉。

## 5 小组分工

**孙旭磊**：组织策划、前端与前后端衔接、数据处理 API 编写、公式插入 API 编写、数据处理示例程序编写、帮助文档编写、代码审查、海报和宣传片制作、实验报告修订，5 个实验处理模块的开发；**张学涵**：实验报告编写、数据处理 API 修订，5 个实验处理模块的开发；**秦沁**：宣传推广、实验报告修订，4 个实验处理模块的开发；**赵奕**：技术与安全支持、实验报告修订，4 个实验处理模块的开发；**鲍政廷**：海报制作、实验报告修订，6 个实验处理模块的开发。具体的实验模块分工可参考本项目源码中的 README.md。

## 6 总结与收获

本次实践不仅能帮助其他同学更轻松地完成大物实验报告，我们自己也受益良多。

**分工与合作** 我们分工明确，每个人的任务都有截止时间，这使我们小组的进度有序推进。以往的经历中，代码与相关工作往往都是独立完成，代码规范与项目进程完全由自己安排。但是在这种大工程中，相关代码需要符合规

范，需要与队友交接，工作进度也要与队友进度相符。在这种分工体系下，每个人都要完成自己的任务，并顾及与他人的交互。

**代码规范性** 本次大作业中，我们建立了统一的码风，并制定了自主编写的 API 的使用说明。这样做一方面可以使得产品最终具有一致性——不同人写的代码能够基本一致；另一方面也使得最终的检验与调整能够更加方便——规范的代码提高了代码的可读性，降低了代码的审核成本。

**软件开发技巧** 在本次实践中，我们使用 git 进行协作，代码注释清楚，帮助文档详细。这大大提高了我们的开发效率。

接下来，我们计划将本项目开源，并增添诸如“网页上输入数据”“手写实验数据 OCR”等功能。更进一步地，我们会考虑支持二级大物实验，以及将本产品推广至其他院校。

## 7 参考资料

Matplotlib, pandas, SciPy, SymPy, python-docx 等包的使用方法参考了官方文档，遇到的各种问题则是查阅 Stack Overflow 等网站上的讨论。在 Word 中渲染 L<sup>A</sup>T<sub>E</sub>X 公式的方法，参考了 python-docx 开源仓库里一份 Github Issues 中提出的[解决方案](#)。

另附

- 大雾实验工具网址：<https://dawu.feixu.site/>
- 宣传片地址（发布于 B 站）：<https://www.bilibili.com/video/BV193411G7aP>
- 项目源代码（文件目录见其中的 README.md）