

选题



掘金已支持深色模式

在“我的设置-通用设置”中即可配置

[点击前往](#)

好的选题和内容切入点是一本小册成功的基础，但同时也是最难的地方。要兼顾内容的难度和受众大小，并从中找出有差异性的角度是我们思考小册选题的基础。

举例来说，Git 基本语法的内容难度不高、但受众很大，因而就需要作者能够想出好的内容差异性；高并发架构设计的内容难度高、但受众有限，差异性依赖于作者是否有特别的实践经验；程序的设计模式是一个有难度的内容，而且大多数开发者如果想要提升自己的能力都需要学习，这个品类就很有可能出现大家需要的小册。

当然，作者并不是一个人在战斗！掘金的团队非常愿意根据掘金产品上的数据和我们的社区运营经验，来和每位作者一起分析当下的开发者最需要的内容。

! 复制代码

- 1 掘金小册的选题原则：以一个明确的开发者需要的问题为出发点

👉 欠佳例子

TensorFlow 入门

- 问题：小册的选题问题不明确。入门？入门到怎样的水平？要能解决怎样的问题？是要把 TensorFlow 全都讲一遍？是要服务于什么具体的工作需求？要讲多细？这个题目没有办法把这些需求明确的表达清楚。
- 👍 选题优化：
 - TensorFlow 基本语法和模型运行指南
 - MNIST 机器学习方法及 TensorFlow 的运行实现
 - 基于 TensorFlow 实现一个图片自动分类器

Vue.js 2.0 升级踩坑记

- 问题：小册的选题过于细节。文段内容不足以支撑小册的产品模式，更适用于文章或是问答形式，且每一个业务的具体需求不同在升级中遇到的问题不尽相同，因而小册凝练出的内容不够有价值。
- 👍 选题优化：



- Vue.js 的前端组件化实践

搞下 ARKit

- 问题：小册的选题过于随意。既无法表达具体问题，标题完整性不足，让读者觉得内容本身质量欠思考、甚至不佳。
- 选题优化：
 - ARKit 快速实现：做一个桌子上的小动物
 - iOS ARKit 组件功能详解



掘金已支持深色模式

在“我的设置-通用设置”中即可配置

[点击前往](#)

确定小册大纲

提纲挈领，抓住要领才能让小册价值得以发挥。因而，在我们选择好了一个合适的选题之后，就需要现对整个小册的内容进行一个拆解，从而确定小册的大纲。

2.4 有用的工具：type、dir 和 help	28
2.4.1 type	28
2.4.2 dir	28
2.4.3 help	28
2.5 综合运用	30
2.6 代码的含义	31
2.7 小结	32
第 3 章 供机器读取的数据	34
3.1 CSV 数据	35
3.1.1 如何导入 CSV 数据	36
3.1.2 将代码保存到文件中并在命令行中运行	36
3.2 JSON 数据	39
3.3 XML 数据	41
3.4 小结	44
第 4 章 处理 Excel 文件	56
4.1 安装 Python 包	58
4.2 解析 Excel 文件	59
4.3 开始解析	60
4.4 小结	71
第 5 章 处理 PDF 文件，以及用 Python 解决问题	73
5.1 尽量不要用 PDF	75
5.2 解析 PDF 的编程方法	74
5.2.1 利用 slate 库打开并读取 PDF	75
5.2.2 将 PDF 转换成文本	77
5.3 利用 pdfminer 解析 PDF	78
5.4 学习解决问题的方法	92
5.4.1 练习：使用表格提取，接用另一个库	94
5.4.2 练习：手动清洗数据	98
5.4.3 练习：试用另一种工具	98
5.5 不常见的文件类型	101
5.6 小结	101
第 6 章 数据获取与存储	103
6.1 并非所有数据生而平等	103
6.2 真实性核查	104
6.3 数据可靠性、数据清洁度和数据寿命	105
6.4 寻找数据	105
6.4.1 打电话	105
6.4.2 美国政府数据	106
6.4.3 全球政府和城市开放数据	107
6.4.4 组织数据和非政府组织数据	109
6.4.5 教育数据和大学数据	109
6.4.6 医学数据和科学数据	110
6.4.7 众包数据和 API	111
6.5 案例研究：数据调查实例	111
6.5.1 埃博拉病毒危机	111
6.5.2 列车安全	112
6.5.3 足球运动员的薪水	112
6.5.4 重工业	113
6.6 数据存储	113
6.7 数据库简介	114
6.7.1 关系型数据库：MySQL 和 PostgreSQL	116
6.7.2 非关系型数据库：NoSQL	117
6.7.3 用 Python 创建本地数据库	118
6.8 使用简单文件	118
6.8.1 云存储和 Python	119
6.8.2 本地存储和 Python	119
6.9 其他数据存储方式	119
6.10 小结	121
第 7 章 数据清洗：研究、匹配与格式化	121
7.1 为什么要清洗数据	122
7.2 数据清洗基础知识	123
7.2.1 找出需要清洗的数据	131
7.2.2 数据格式化	135
7.2.3 找出离群值和不良数据	140
7.2.4 找出重复值	143
7.2.5 模糊匹配	146
7.2.6 正则表达式匹配	150
7.2.7 如何处理重复记录	151
7.3 小结	153
第 8 章 数据清洗：标准化和脚本化	153
8.1 数据归一化和标准化	154
8.2 数据存储	156
8.3 找到适合项目的数据清洗方法	157
8.4 数据清洗脚本化	170
8.5 用新数据测试	172
8.6 小结	172

目录 | ix

@稀土掘金技术社区

与市面上的书籍章节不同，掘金小册只支持第一级目录，我们称之为**小节**。这背后的目的是希望在掘金小册的内容大纲足够干练精准，减少读者获取信息的成本。章节上篇幅亦不要过于冗长，小节呈现一个有体系、有逻辑性的顺序，从而完成整个小册的小节结构。



1. 一本书中包含大量线上可以找到且更新更及时的内容
2. 一本书冗长的内容中有很多章节读者不知道是否和自己的需求直接相关

注意，我这里绝对没有说一本完整的书籍不好，完整性和全面性重要的部分。但是，我们面对的技术市场是一个高速迭代的市需要快速理解一个新技术的优势和具体实践；而作者也要应大多数快速学习的内容，进而再持续深入。



掘金已支持深色模式

在“我的设置-通用设置”中即可配置

[点击前往](#)

几个掘金小册的推荐大纲结构

一、开发实现顺序

许多技术类内容都是要顺着具体的开发引导来的，因而按照实现的顺序一层一层、一步一步地讲解，是可以帮助读者从头到尾把内容连贯起来的。

以[如何写一本掘金小册](#)为例：先讲掘金小册是什么，然后读者就会问那怎么用啊？然后就会想要写什么怎么写？然后就是要如何售卖？上线之后有了读者就会思考如何来维护和讨论？有了读者获得了内容就可能出现版权问题？最后，讲明白掘金小册的需求，就希望更多人来写，因而留下申请成为小册作者的方法。

- 第一部分：介绍我们要做什么
- 第二部分：按照实现顺序按部就班地详解流程
- 第三部分：完成的结果，照应主题，留下读者可以继续深入学习、执行的方法

二、概念到细节

很多技术内容的讲解，尤其是一些新的技术，最先要讲明白背后的概念，让读者对这个内容有了一个概括性的了解，然后再深入背后的细节完成实体的内容。

以[Git 原理详解及实用指南](#)为例：Git 虽然为大家熟悉，但它本身是一个分布式版本管理工具（DVCS）及其基本的概念需要让读者先理解。然后来讲一些核心概念，如 repository、branch、stage 等等；然后讲基本操作以及最后的一些工作中会遇到的实用方法。一步步从概念走到细节。

- 第一部分：技术的基本概念
- 第二部分：技术的核心概念
- 第三部分：技术的基本功能和实用方法
- 第四部分：技术的高级实用技巧
- 第五部分：总结 + 深入学习的资源列表



还有一种小册的大纲结构是以经验总结、最佳头践、问题解决方案为主的平行内容。这样往往一开始先讲明白小册的核心目的和价值，然后就会围绕主题一个一个讲解解决方法、或者一个个的补充内容。

以[响应式编程 —— RxJava 高阶指南](#)为例：小册的目标是去帮助 RxJava 的高阶问题。在第一节解释清楚 Rx 的价值之后就会去帮助 RxJava 的高阶问题，但是各个问题之间的关系相对是平行的。

掘金已支持深色模式

在“我的设置-通用设置”中即可配置



[点击前往](#)

- 第一部分：小册要去解决的问题的重要性的核心内容
- 第二部分：一个一个地讲解内容
- 第三部分：总结

开头和结尾的重要性

掘金小册的小节作者可以开放试读功能，而一般第一小节都会选择试读。因而，一本小册的第一小节往往是读者对小册的第一印象。

我们建议一个小册的开头：

- 清楚讲解小册的核心价值和目的
- 配以合理的介绍性的文字、图片、代码，甚至是 Demo
- 对整本小册的结构大纲有一个概述性的介绍，为后面的核心内容做铺垫和准备

读者完成一本小册的阅读，到达最后一个小节，可以获得其期待的知识信息。但是，读者往往在这一刹那希望可以获得更多、可以与作者和其他读者交流、可以有更多的资源和技术可以更深入的学习下去。

因而我们建议一个小册的结尾：

- 回顾整本小册，对内容、核心产出、重点做一次总结
- 援引一些资源、开源库、书籍、文章链接，让读者可以继续学习
- 给读者以与作者继续保持联系的方法

组织小册内容



[! 复制代码](#)



3. 小节之间应连贯有逻辑，遵从一个大网排布规则
4. 每一小节的平均阅读时长在 10 分钟左右
5. 每一个小节都应该在文段里有一个标题
6. 每一个小节最后要有一个小结来总结本小节的核心内容
7. 可以根据小册的特点，适当地设置一些练习题或问答题
8. 最后一个小节应该有一个总结和延伸阅读的入口
9. 图片要保证无版权风险，且大小合宜、描述清晰
10. 准确的代码区块，切忌大段粘贴代码
11. 内容排版中善用 `h2`、`h3`、`blockquote`、`code`、`ul`、`ol` 等元素
12. 在内容上多使用链接去援引外部内容



掘金已支持深色模式
在“我的设置-通用设置”中即可配置

[点击前往](#)

掘金责任编辑

为了帮助作者选题、组织内容和运营推广，掘金团队会为每位掘金小册的作者配一位责任编辑，他会与作者一起完成一本小册从最初内容思考到最终上线出售的完整过程。

申请成为小册作者

留言

看完啦，[立刻购买](#) 分享一下感受吧~

全部评论 (26)

团长在此

PHP 1年前

打卡

点赞 回复 ...

一个被摄影耽误的...

Android Bug制造机 1年前

怎么成为小册作者啊

点赞 1 ...

掘金小册 1年前

用户6137690653... 前端开发工程师 @ 正... 1年前

干了这么多年开发，是时候输出一些有意义的内容了，帮助更多开

👍 3 🗨 回复 ...

掘金已支持深色模式

在“我的设置-通用设置”中即可配置



[点击前往](#)

flightmakers 最酷的Maker @ 掘金&... 2年前

同问 #起飞计划

👍 点赞 🗨 回复 ...

春水煎茶_ 前端工程师 @ 阿狸88 2年前

为啥评论都重复两遍啊

👍 1 🗨 回复 ...

ming酱71471 前端工程师 @ 互联网... 4年前

怎么成为小册作者啊

👍 点赞 🗨 回复 ...

ming酱71471 前端工程师 @ 互联网... 4年前

怎么成为小册作者啊

👍 点赞 🗨 回复 ...

上车兄弟 4年前

oracle不在是个梦

👍 1 🗨 回复 ...

上车兄弟 4年前

oracle不在是个梦

👍 点赞 🗨 回复 ...

Andrewpqc Backend Developer. ... 4年前

搞下 ARKit, 真的有人投过这个标题吗? 太搞笑了

👍 点赞 🗨 1 ...

皮卡qiu 4年前

还没。

👍 点赞 🗨 回复 ...



👍 点赞 💬 1 ...

皮卡qiu 4年前

还没。

👍 点赞 💬 回复 ...



掘金已支持深色模式

在“我的设置-通用设置”中即可配置

[点击前往](#)

沉默王二 微信搜 @ 沉默王二 4年前

欠佳例子：Java的那些事儿

👍 点赞 💬 回复 ...

沉默王二 微信搜 @ 沉默王二 4年前

欠佳例子：Java的那些事儿

👍 点赞 💬 回复 ...

浪里行舟 前端 @ 联系微信frontJS 4年前

请问小册可以包括已经发布文章？还是每一篇都要首发

👍 点赞 💬 1 ...

皮卡qiu 4年前

可以包括，评判看文章质量。

👍 点赞 💬 回复 ...

浪里行舟 前端 @ 联系微信frontJS 4年前

请问小册可以包括已经发布文章？还是每一篇都要首发

👍 点赞 💬 1 ...

皮卡qiu 4年前

可以包括，评判看文章质量。

👍 点赞 💬 回复 ...

月翎魂雨 web前端 @ nio 4年前

我试试

👍 点赞 💬 回复 ...

 点赞  回复 ...

碧青Johnny

前端工程师 @ 字节跳动 5年前

“开头和结尾的重要性”一节有个笔误哈，对整个小节（->“册”）

 点赞  1 ...



掘金已支持深色模式

在“我的设置-通用设置”中即可配置

[点击前往](#)

阴明（作者） 5年前

[谢谢反馈](#)

 点赞  回复 ...

碧青Johnny

前端工程师 @ 字节跳动 5年前

“开头和结尾的重要性”一节有个笔误哈，对整个小节（->“册”）的结构大纲有一个概述性的介绍

 点赞  1 ...

阴明（作者） 5年前

[谢谢反馈](#)

 点赞  回复 ...