基于开源编辑器的协同编辑方案

Peter

August 16, 2024

自我介绍

- 全栈工程师
- 一名“老鹅”（曾就职于鹅厂腾讯）
- 创业者，TeXPage 创始人

开源编辑器

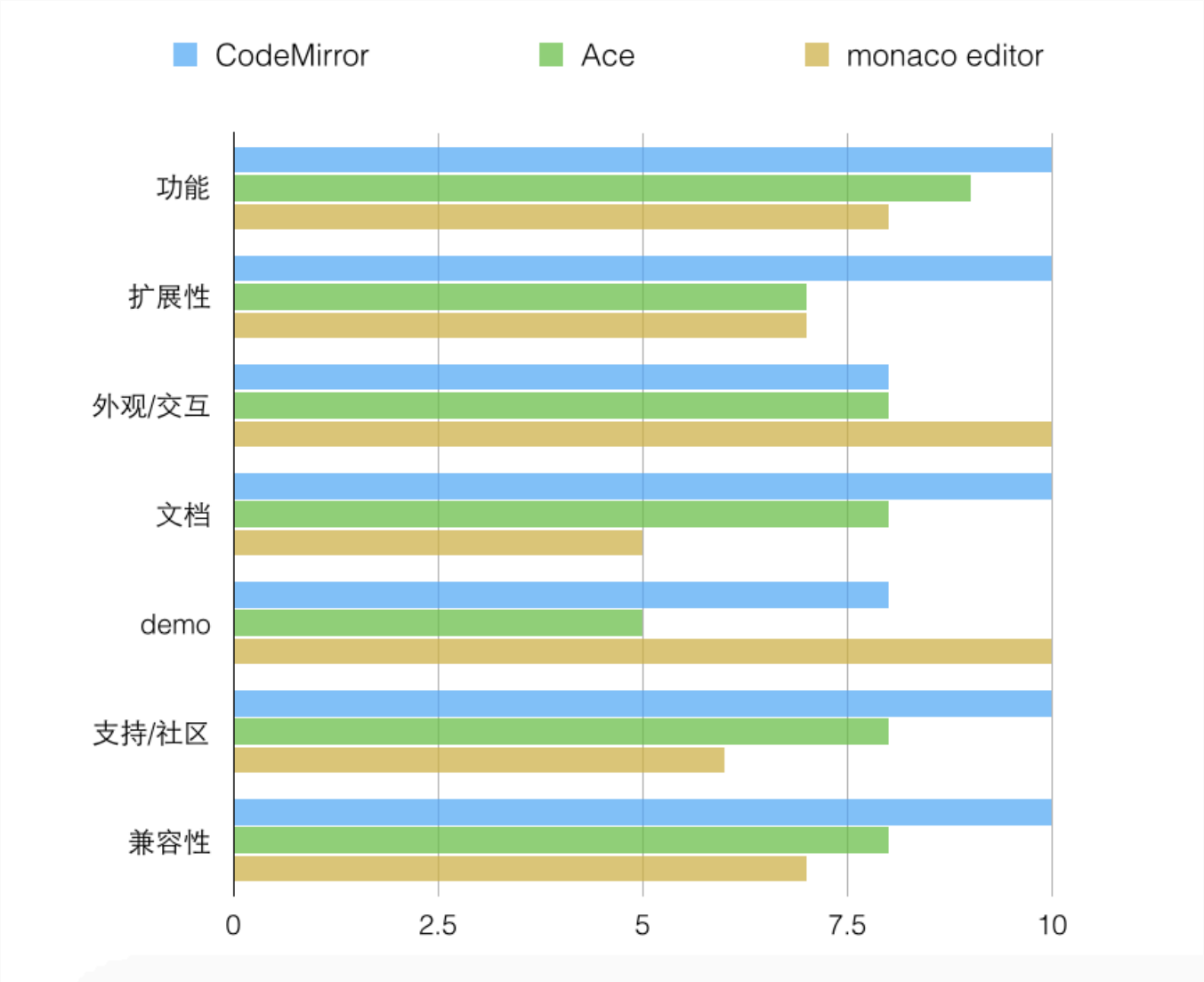
开源编辑器分类

- **Code Editor**
  - CodeMirror
  - Monaco Editor
  - ACE Editor
- **Rich-Text Editor**
  - ProseMirror
  - Quill
  - Draft.js
  - Slate.js (WangEditor)
  - Editor.js

开源编辑器的核心模块

- **文本编辑**
  - 使用隐藏默认样式的 textarea，需要模拟光标和选中状态
  - contenteditable，光标和选中状态利用浏览器的原生支持
- **语法分析和主题**
  - CodeMirror - Lezer
  - Monaco - Monarch
  - ACE
- **数据模型和渲染**
- **自定义扩展**

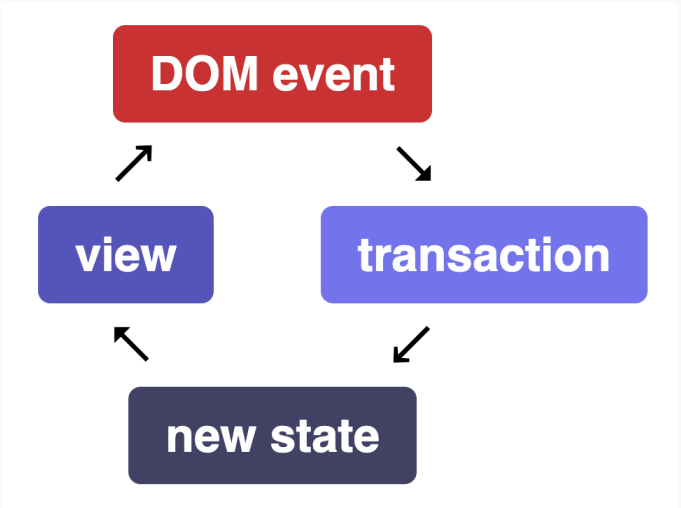
Code Editor 对比分析



from: https://sq.sf.163.com/blog/article/184733100361850880

CodeMirror 核心特点

- **核心逻辑强悍**，非核心逻辑略显粗糙
- **模块化**，任何一个功能模块都可以自由定制和替换
- 主逻辑均是**纯函数**，没有副作用



开源编辑器的选型

基本使用：
都很优秀，选哪个不会有太大的差别

做产品：

- 关注核心功能，功能丰富程度不太重要
- API 丰富度
- 熟悉源码，文档可能会误导，代码不会

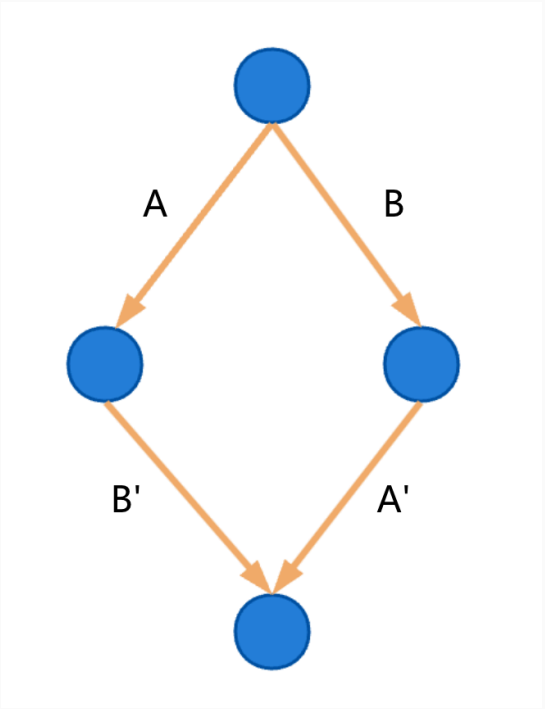
协同编辑

协同编辑算法

- **OT**（Operational Transformation）
  - Google Docs
  - 腾讯文档
- **CRDT**（Conflict-free Replicated Data Type）
  - Notion
  - Figma
  - Redis 处理分布式系统中的数据一致性问题

OT

- 技术方案相对成熟
- 服务端协调所有操作，按特定顺序合并
- 数据结构简单，不易扩展
- 开源库：ot.js

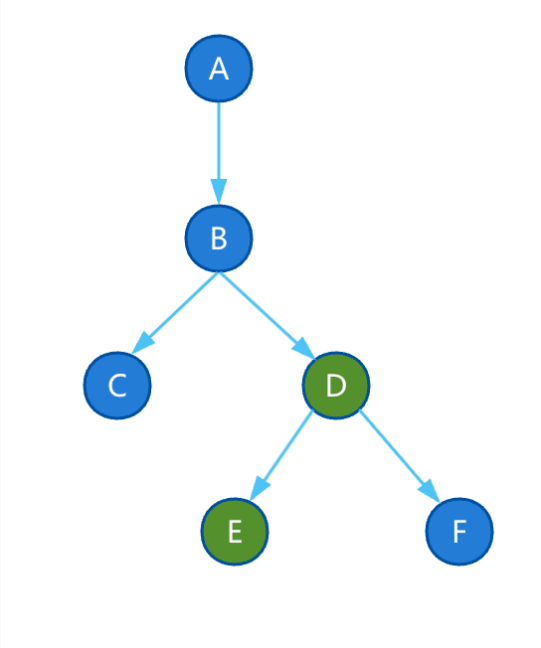


OT 算法实现审阅功能的不足

- OT 算法的数据只记录“当前”状态下的编辑数据
- 难以追踪所有协作者的编辑记录
- 需要大量的定制化开发工作

CRDT

- 分布式，不依赖服务端协调
- 数据结构灵活，易扩展
- 存储开销大
- 开源库：Yjs

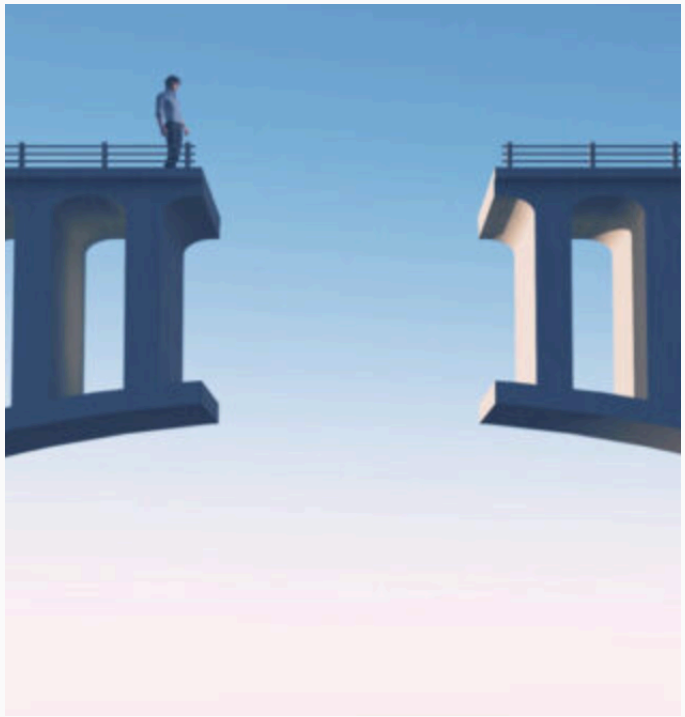


CRDT 数据压缩带来的挑战

- 压缩数据后引起的**编辑中断**
- 压缩数据后引起的**数据冲突**
- 部分编辑**数据丢失**

工程实践

- **协同编辑算法和编辑器适配**
- **优化并发协同编辑数据处理速度**
  - 实时处理核心协同编辑
  - 异步处理非核心逻辑
- **边界情况处理**
  - 网络中断与重连后的数据同步
  - 超大规模的粘贴
  - ...
- **测试难度大**



Thanks !