**关于Rust 语言.docx文档格式解析的调研报告**

*这是一份Rust生成的关于Rust语言.docx文档格式解析的调研报告。*

1. **背景**

这份调研起始于钉钉面试后的一份作业，要求读取一份docx文档并打印到屏幕。我通过检索Crates.io查找了Rust语言下docx文档格式解析的库，对当前可供使用或参考的解析库做了比较。

1. **可用的Rust库比较(till 2020-10-21 21:00)**

|  |  |  |  |
| --- | --- | --- | --- |
| **features/crate** | **docx-rs** | **docx** | **dotext** |
| pure rust | yes | yes | yes |
| stars on github | 57 | 28 | 18 |
| last version in crates.io | 0.2.9 | 1.1.2 | 0.1.1 |
| last update in crates.io | 9 days ago | 6 months ago | 3 years ago |
| last update in github | 9 hours ago | 6 months ago | 9 months ago |
| wasm support | yes | not clear | not clear |
| docx reading | yes | yes | only read to string |
| WordprocessingML parsing | yes | yes, partial | no |
| docx writing | yes | yes | no |
| documentations | no doc | little | no |
| other format parsing | no | no | support docx/xlsx/odt/ods/pptx |
| raw xml parser | xml-rs | strong-xml | quick-xml |

这几个库目前文档都不全，docx-rs 正在快速迭代开发中，文档也会逐渐丰富，支持WASM，是目前Rust语言docx解析库的首选。如果要开发一个新的文档格式解析器，其也可以作为一个很好的参考。其他两个库的更新频率较低。另外，这三个库选择不同的XML解析库，编写解析库时建议进行性能评估。

docx-rs 目前尚不支持Footer, TableContents 等一些WordprocessingML的元素，对space等属性的支持也不好，在使用中需要注意。

1. **Rust示例代码**

当前文档存储在我的GitHub中，克隆仓库后使用`cargo run`即可生成在当前目录下。详情访问：https://github.com/zitsen/docx-demo.rs