Excelize 开源基础库助力企业信息系统协同创新

Excelize Open-source Library Empowered of Enterprise IT Systems Collaborative Innovation

续日

XU RI

阿里巴 技术专家, Excelize 作者 Alibaba Group Technical Specialist, author of Excelize







续日

GitHub: @xuri

https://xuri.me

软件工程师, Excelize 开源基础库作者 阿里巴巴技术专家 曾就职于百度、360等公司 前百度 Go 语言编程委员会成员, 从事百度 Go 语言研发体系的建设工作





内容提要

业务背景:企业级应用中,办公文档数据协同的难点

应用案例: Excelize 如何助力企业信息系统的协同创新

架构设计:办公文档格式国际标准实现过程中面临的挑战的解法

总结展望:Excelize 开源数据洞察、WebAssembly 支持和未来规划





业务背景: 电子表格办公文档的应用领域广泛

日历&计划

日程管理 计划任务 轮班时间表 名册模版 个人计划 每日代办

项目管理

甘特图 项目规划 敏捷看板 关键路径分析 代办列表 代外列表 时间线模版 工作项拆解

预算 & 财务

人力资源 & 工资表

员工工资 考勤记录 开支报告 报销表格 假期追踪 面试问题和反馈 人力资源知识库 员工满意度调研

销售

商机管理 CRM 内部销售 销售渠道管里 客户评价追踪 会员名录 销售业绩管理

研发管理

缺陷追踪 用户调研 价值风险 客户反馈 技术路线图 项目进展追踪 用户增长实验记录 产品规划表

内容生产

视频创作 播客编辑日历 媒体编辑日历 选角表和试镜表 剧组人员管理 内容营销管理

更多领域 & 行业

金融 医疗教育 能源 交通 制造业 服务业





企业办公文档数据协同场景

由于企业的业务特点,难以实现一套系统满足企业全场景应用

大量场景需要企业应用集成 (EAI: Enterprise Application Integration)

办公文档如同企业的"数字化细胞",成为系统间数据协同的一种常用媒介







Excelize: 开源电子表格办公文档基础库

开源代码:完全开源的电子表格办公文档基础库 https://github.com/xuri/excelize

开源合规: 采用 BSD 3-Clause "New"or "Revised" License 开源许可协议

开放标准: 基于 ECMA-376, ISO/IEC 29500 国际标准实现

互操作性: 支持读写 Excel、WPS、OpenOffice 等办公应用创建的电子表格文档

文档格式: XLAM / XLSM / XLSX / XLTM / XLTX 多种办公文档格式支持

主要特点:兼容性好、高性能、安全、功能丰富、支持跨平台





案例1: Excelize 在飞书企业协作与管理平台中的应用

借助 Excelize 基础库解析和生成电子表格中的数据,实现批量对员工信息的导入和导出,帮助组织人员信息在线化



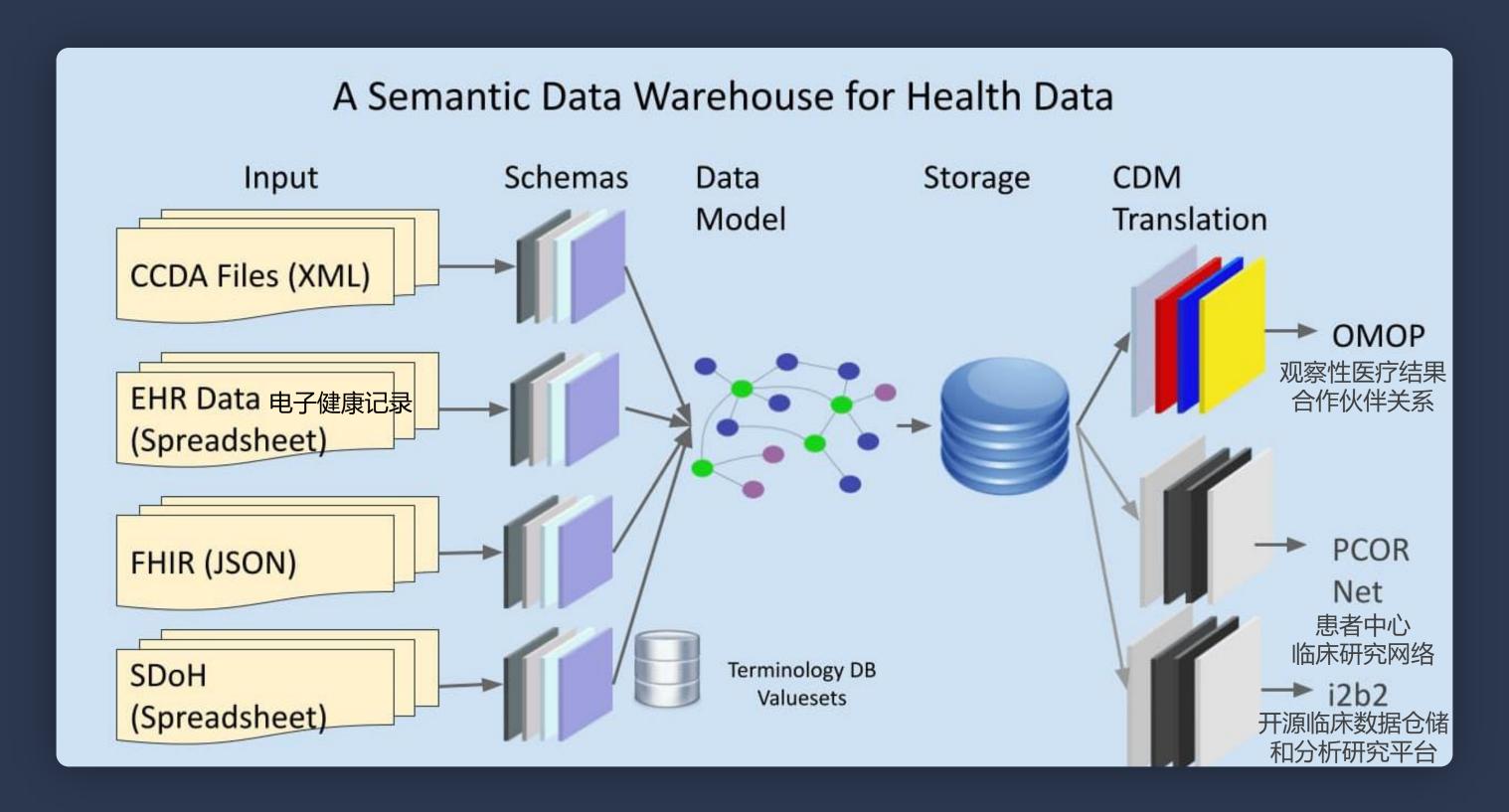
https://juejin.cn/post/7117578224854368263





案例2: Excelize 在医疗信息系统中的应用

ONC 在可分层医疗信息数据处理框架中,通过集成 Excelize 基础库对 EHR 和 SDoH 电子表格数据格式进行解析,实现异构数据源转换协调,并与 OMOP、i2b2 等系统连接协同



https://github.com/onc-healthit/2021-LEAP-Semantic-Interoperability





应用1:读写电子表格文档

创建工作簿、设置单元格的值

```
f := excelize.NewFile()
// 创建一个工作表
index, _ := f.NewSheet("Sheet2")
// 设置单元格的值
f.SetCellValue("Sheet2", "A2", "Hello, world!")
f.SetCellValue("Sheet1", "B2", 100)
// 获取工作表中指定单元格的值
cell, err := f.GetCellValue("Sheet1", "B2")
if err != nil {
   fmt.Println(err)
   return
fmt.Println(cell)
// 设置工作簿的默认工作表
f.SetActiveSheet(index)
// 根据指定路径保存文件
if err := f.SaveAs("Book1.xlsx"); err != nil {
   fmt.Println(err)
```

打开工作簿、读取工作表中的单元格

```
f, err := excelize.OpenFile("Book1.xlsx")
if err != nil {
    fmt.Println(err)
    return
defer func() { // 关闭工作簿
    if err := f.Close(); err != nil {
        fmt.Println(err)
}()
rows, err := f.GetRows("Sheet1")
if err != nil {
    fmt.Println(err)
for _, row := range rows {
    for _, colCell := range row {
        fmt.Print(colCell, "\t")
```





应用2:数据透视分析

借助 Excelize 创建数据透视表按月对各区域在售的商品销售额进行分类汇总,并支持按销售区域、时间和分类进行筛选



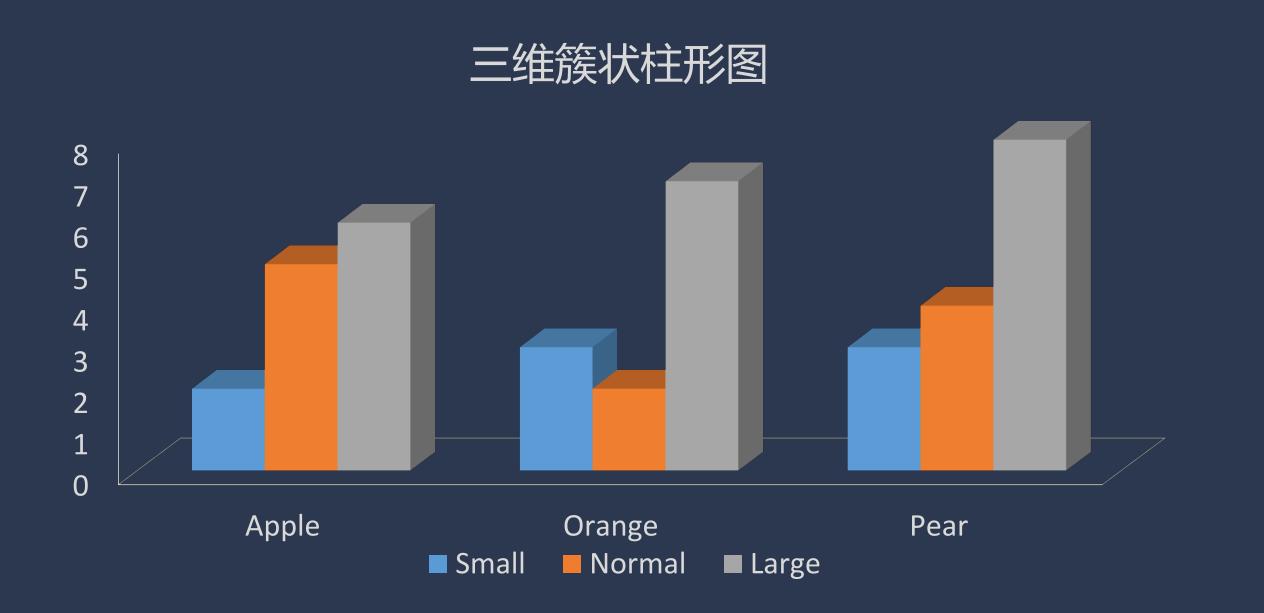
```
f.AddPivotTable(&excelize.PivotTableOption{
                    "Sheet1!$A$1:$E$31",
   DataRange:
   PivotTableRange: "Sheet1!$G$2:$M$24",
    Rows: []excelize.PivotTableField{
        {Data: "Month",
        DefaultSubtotal: true},
        {Data: "Year"}},
   Filter: []excelize.PivotTableField{
        {Data: "Region"}},
    Columns: []excelize.PivotTableField{
        {Data: "Type", DefaultSubtotal: true}},
    Data: []excelize.PivotTableField{
        {Data: "Sales", Name: "销售额汇总",
    Subtotal: "Sum"}},
   RowGrandTotals: true,
    ColGrandTotals: true,
   ShowDrill:
                    true,
   ShowRowHeaders: true,
   ShowColHeaders: true,
    ShowLastColumn: true,
})
```





应用3:数据可视化分析

Excelize 支持二维、三维、面积图、条形、圆锥、棱锥、柱形、堆积柱形、圆环图、折线图、饼图、雷达图、散点图、曲面图、气泡图等累计 55 种原生图表与组合图表的创建



```
f.AddChart("Sheet1", "E1", &excelize.Chart{
    Type: excelize.Col3DClustered,
    Series: []excelize.ChartSeries{
                        "Sheet1!$A$2",
            Name:
            Categories: "Sheet1!$B$1:$D$1",
            Values:
                        "Sheet1!$B$2:$D$2",
                        "Sheet1!$A$3",
            Name:
            Categories: "Sheet1!$B$1:$D$1",
                        "Sheet1!$B$3:$D$3",
            Values:
                        "Sheet1!$A$4",
            Name:
            Categories: "Sheet1!$B$1:$D$1",
            Values:
                        "Sheet1!$B$4:$D$4",
    }},
    Title: excelize.ChartTitle{Name: "三维簇状柱形图"},
})
```





架构设计: 相关国际标准(部分)

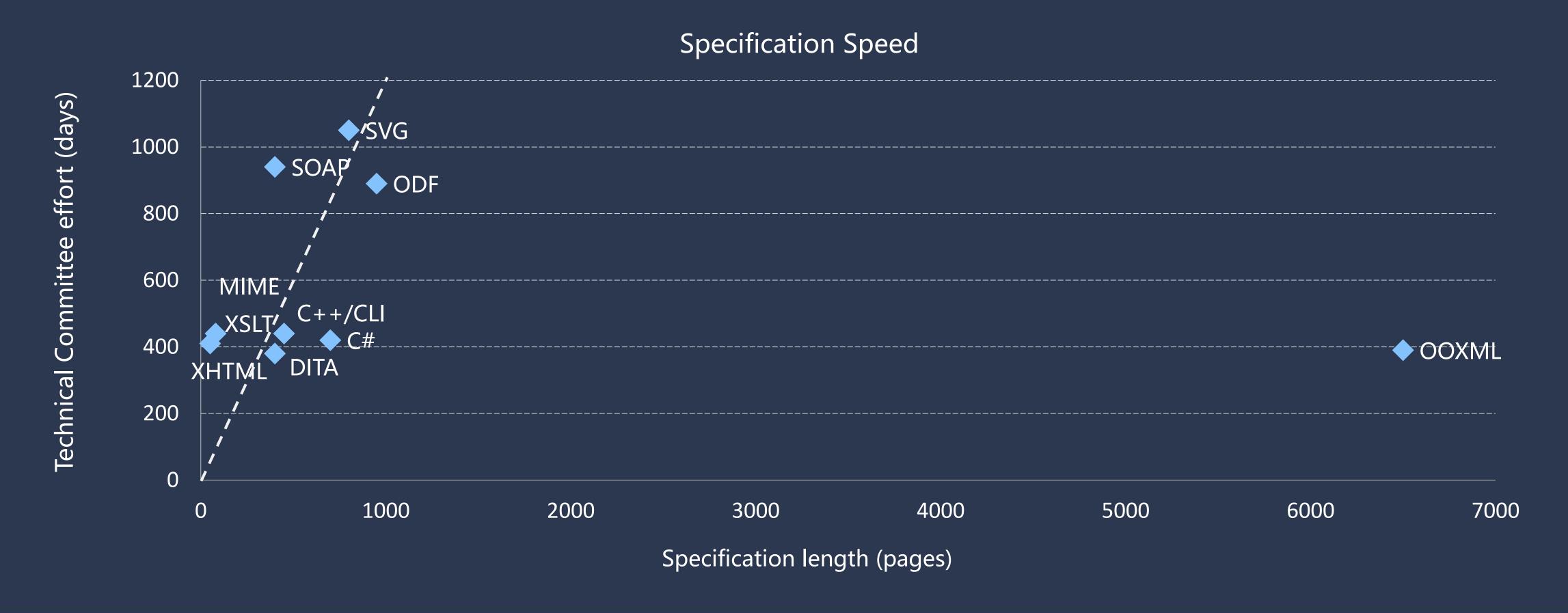
- ISO/IEC 29500 (Office Open XML File Formats)
- ISO/IEC 21320 (Document Container File)
- ISO/IEC 10646-2 (Universal Coded Character Set)
- XML (Extensible Markup Language)
- XSD (W3C XML Schema Definition Language)
- [MS-CFB]: Compound File Binary File Format
- [MS-OFFCRYPTO]: Office Document Cryptography Structure
- [MS-LCID]: Windows Language Code Identifier Reference
- DCMI (Dublin Core Metadata Initiative) Metadata Terms





标准规模对比:庞大复杂的办公文档格式标准体系

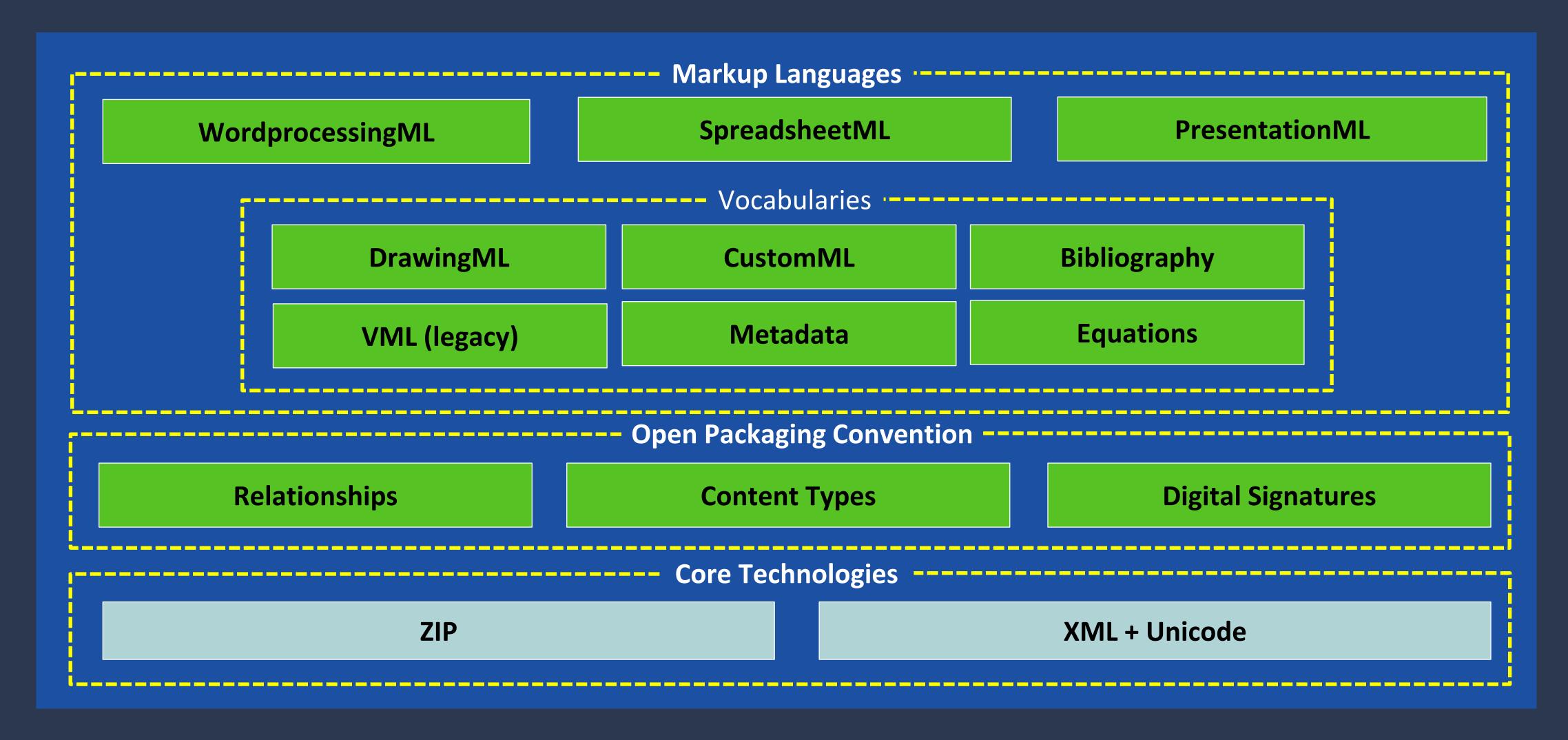
ISO - ISO/IEC 29500-1:2016 ECMA 376 Open Office XML







标准架构

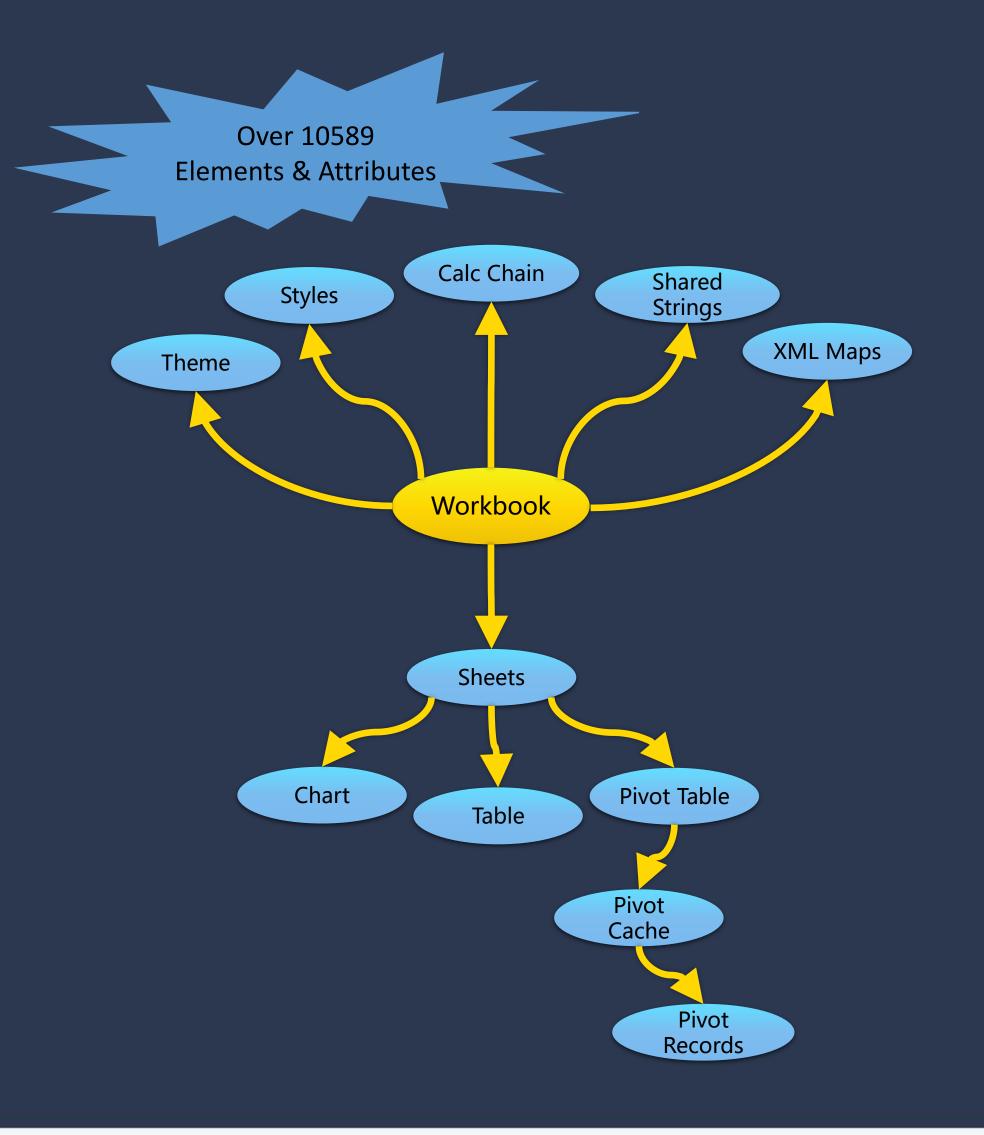






格式解析原理

```
<c r="D3" s="7" t="s">
                                <v> <</v>
        C
                                <xf numFmtId="0" fontId="4" fillId="2" borderId="2" xfId="1" applyBorder="1" />
                                                                                       border>
Revenue
            412.52
                                                                                         <left style=/thick">
         (build-in) font>
                                                                                             <color auto="1"/>
                                                <fill>
                                                    <patternFill patternType="50lid">
                           <sz val="11"/>
                                                                                         </left>
                                                    <fgColor theme="4"/>
                                                                                         <right style="thick">
                           <color_theme="0"/>
                                                                                              color auto="1"/>
                            mame val="Calibri"/>
                                                    <bgColor theme="4"/;</pre>
                                                    </patternFill>
                                                                                          </pight>
                           <scheme val="minor"/>
                                                                                            op style="thick">
                                                </fill>
                       </font>
<a:theme
                                                                                             <color auto="1"/>
    xmlns:a="http://schemas.openxmlformats.org"
                                                                                          </top>
    <a:themeElements>
                                                                                         <bottom style="thick">
        <a:clrScheme name="Office">
                                                                                             <color auto="1"/>
            <a:dk1>
                                                                                         </bottom>
                <a:sysClr val="windowText" />
                                                                                         <diagonal/>
            </a:dk1>
                                                                                     </border>
            <a:lt1>
                <a:sysClr val="window"/
            </a:lt1>
            <a:dk2>
                <a:srgbClr v21="1F497D"/>
            </a:dk2>
            <a:1t2>
                <a:srgbClr val="FAF3E8"/>
            </a:1t2>
            <a:accept1>
                srgbClr val="5C93B4"/>
            </a>:accent1>
<cellStyleX?s count="12">
    <xf //umFmtId="0" fontId="0" fillId="0" borderId="0"/>
    <xf_numFmtId="0" fontId="4" fillId="2" borderId="0" applyBorder="1" applyAlignment="1" applyProtection="1"/>
        </xf>
<cellStyles count="2">
    <cellStyle name="Accent1" xfId="1" builtinId="29"/>
```







Excelize 基础库架构

基础能力

媒体格式支持

元数据解析校验

依赖关系处理

扩展标记处理

样式处理能力

冻结窗格

字体样式

行高 / 列宽

数字格式

色值计算

数据处理能力

数据验证

表单控件

时间处理

Crypto 加解密

单位转换

表格/过滤器

模型处理

模型校验

计算引擎

升级扩展能力

模型纠错验证

字典

富文本

图片/图表

模型组件化

文件格式识别

OPC 封装与解构

2D / 3D 处理

簇状/堆积/面积图

柱形/锥形/饼图

气泡/散点/折线图

属性设置能力

工作簿 / 工作表

可见性设置 行/列处理

属性设置 工作表属性

页眉页脚 视图属性

数据保护 搜索能力

页面布局 流式读写 单元格

数据类型支持

边框样式

选区合并

超链接 批注处理

公式处理

样式索引

计算缓存

单元格样式

数据透视表

条件格式

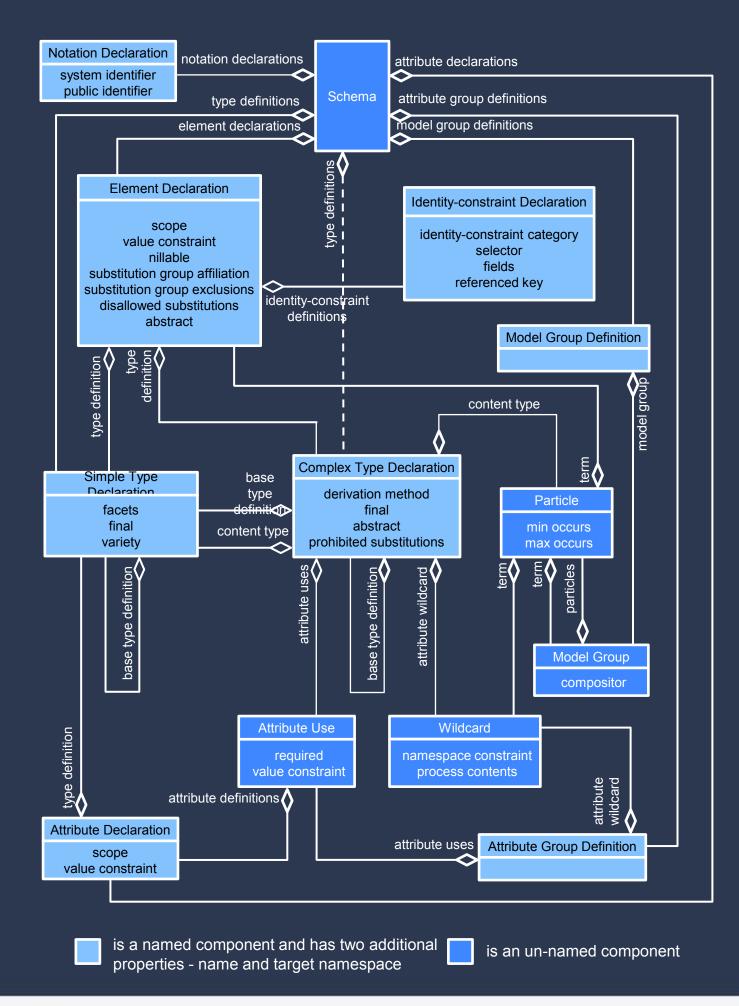
VBA 脚本

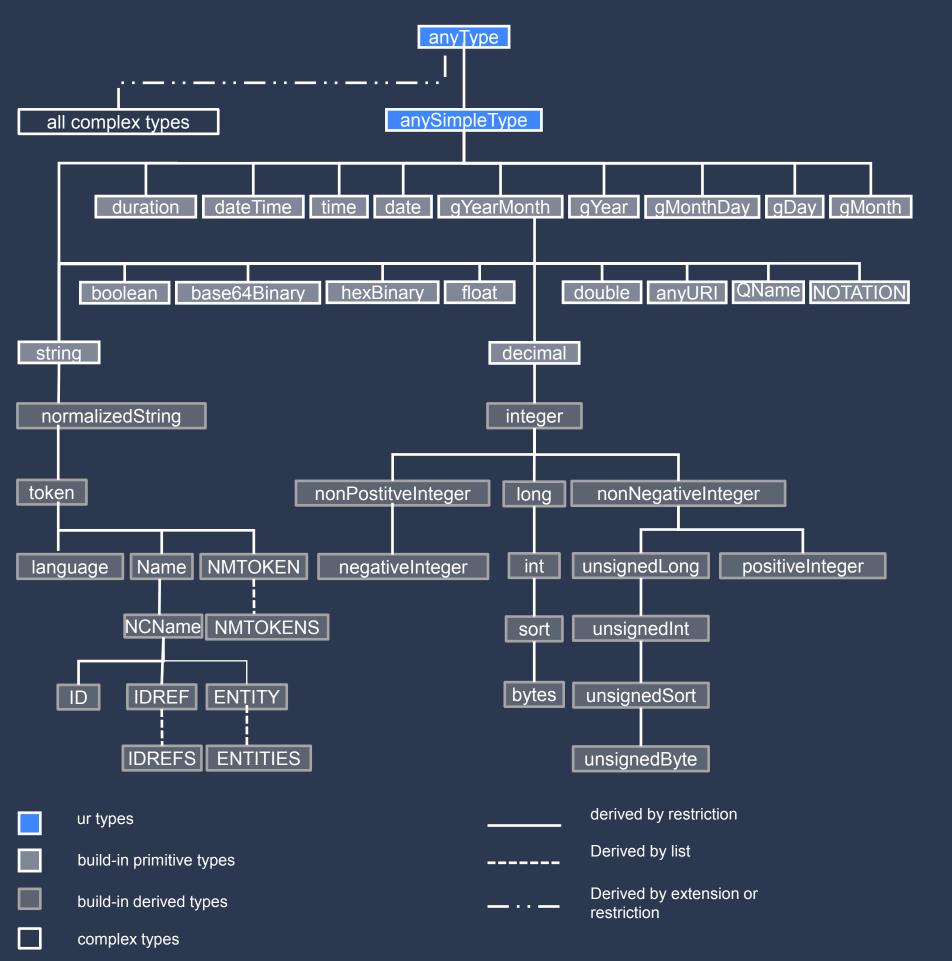




架构设计:兼容性难题如何解决

对文档内存模型的完整度将直接影响文档的兼容性, Excelize 采用基于 XSD 定义的建模

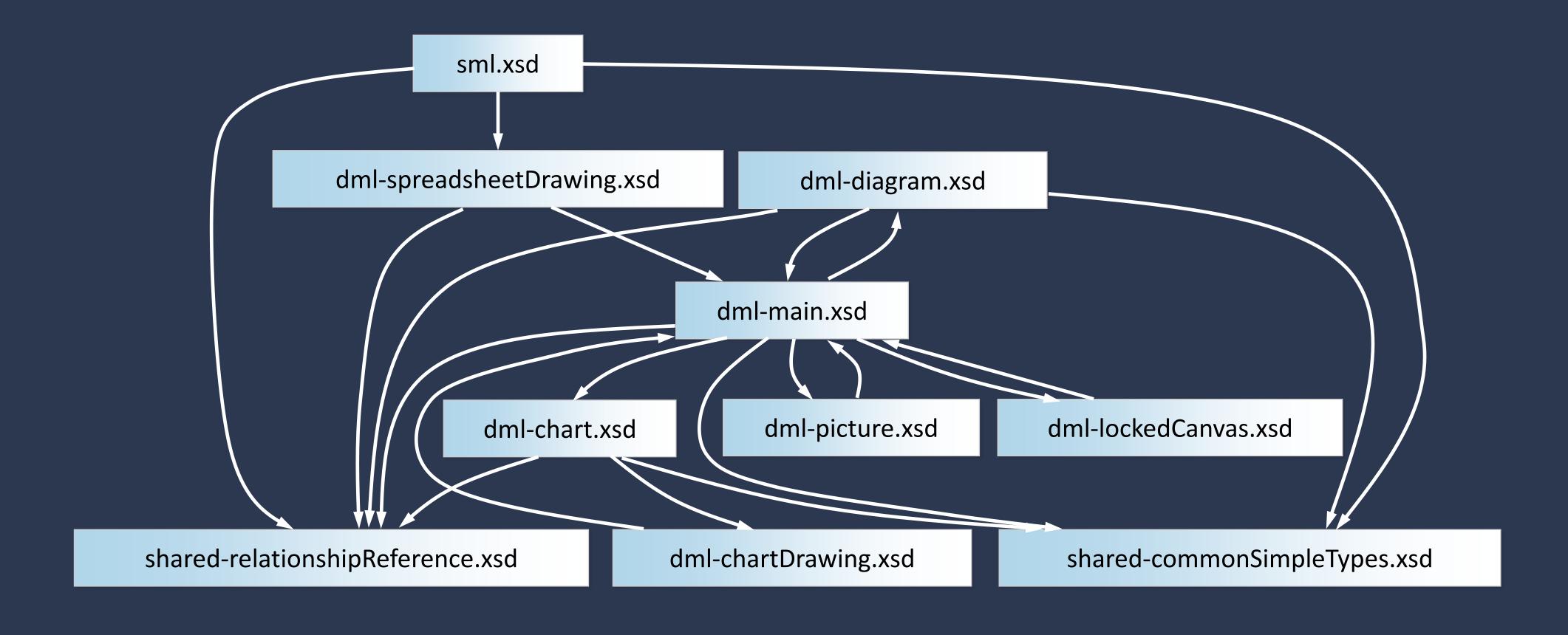








ISO/IEC 29500 中的 XSD 模式的关系

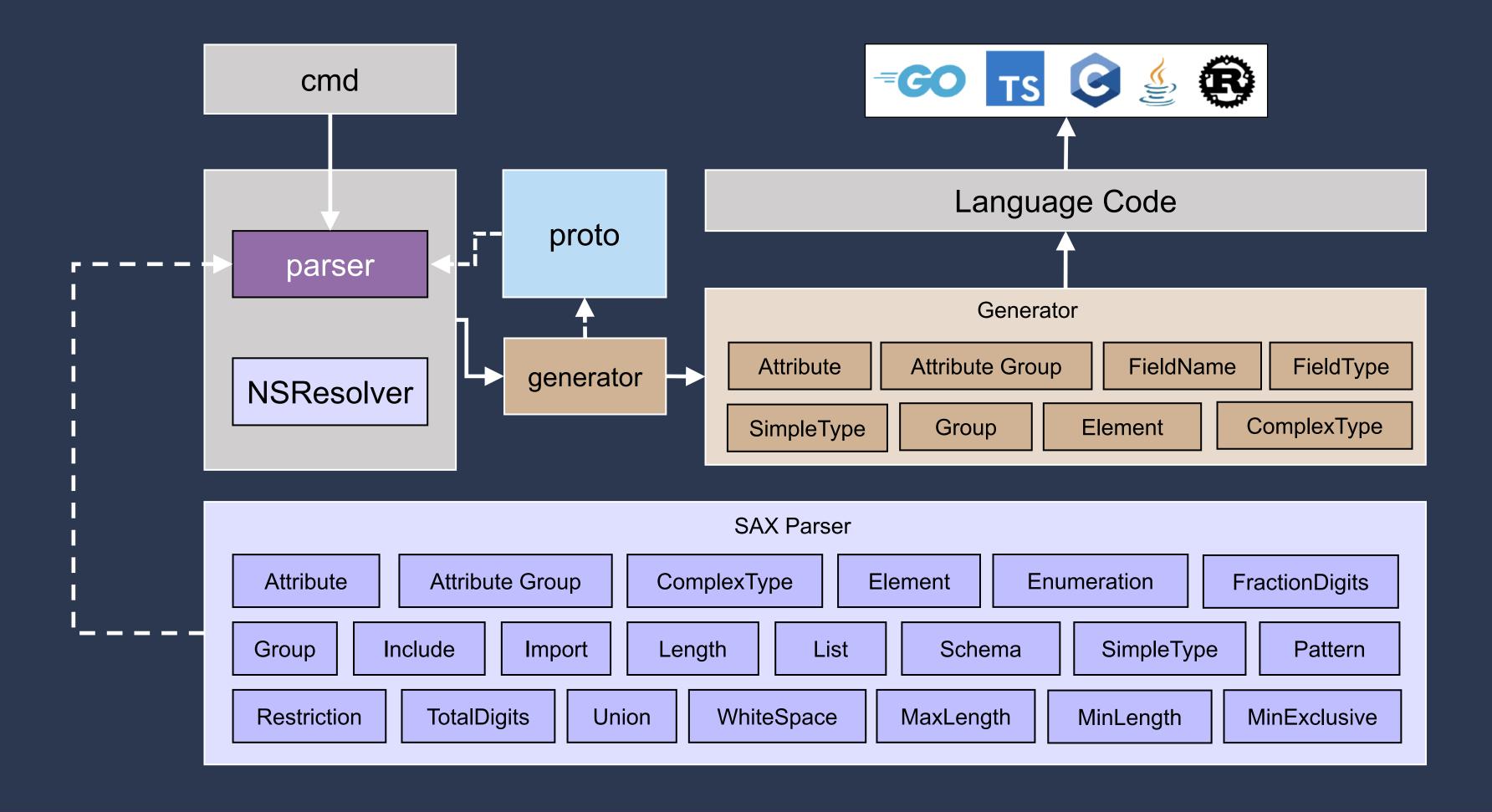






文档模型数据结构代码生成

xgen:基于模式描述定义为强类型语言生成通用语言数据结构代码







xgen: 基于 XSD 的数据结构代码辅助生成



```
安装
```

```
go get -u -v github.com/xuri/xgen/cmd/...
使用
$ xgen -i /path/to/your/xsd -o /path/to/your/output -l Go
$ xgen [<flag> ...] <XSD file or directory> ...
```

- -i <path> Input file path or directory for the XML schema definition
 -o <path> Output file path or directory for the generated code
- -p Specify the package name
- -1 Specify the language of generated code (Go/C/Java/Rust/TypeScript)
- -h Output this help and exit
- -v Output version and exit





示例: 基于 XSD 生成的代码

```
<schema xmlns="http://www.w3.org/2001/XMLSchema" targetNamespace="http://example.org/">
    <simpleType name="myType1">
        <restriction base="base64Binary">
            <length value="10" />
        </restriction>
    </simpleType>
    <complexType name="myType2">
        <simpleContent>
            <extension base="base64Binary">
                <attribute name="length" type="int"/>
            </extension>
        </simpleContent>
    </complexType>
    <complexType name="myType3">
        <simpleContent>
            <extension base="date">
                <attribute name="length" type="int"/>
            </extension>
        </simpleContent>
    </complexType>
    <complexType name="myType4">
        <sequence>
            <element name="title" type="string"/>
            <element name="blob" type="base64Binary"/>
            <element name="timestamp" type="dateTime"/>
        </sequence>
    </complexType>
    <simpleType name="myType5">
        <restriction base="gDay"/>
    </simpleType>
</schema>
```

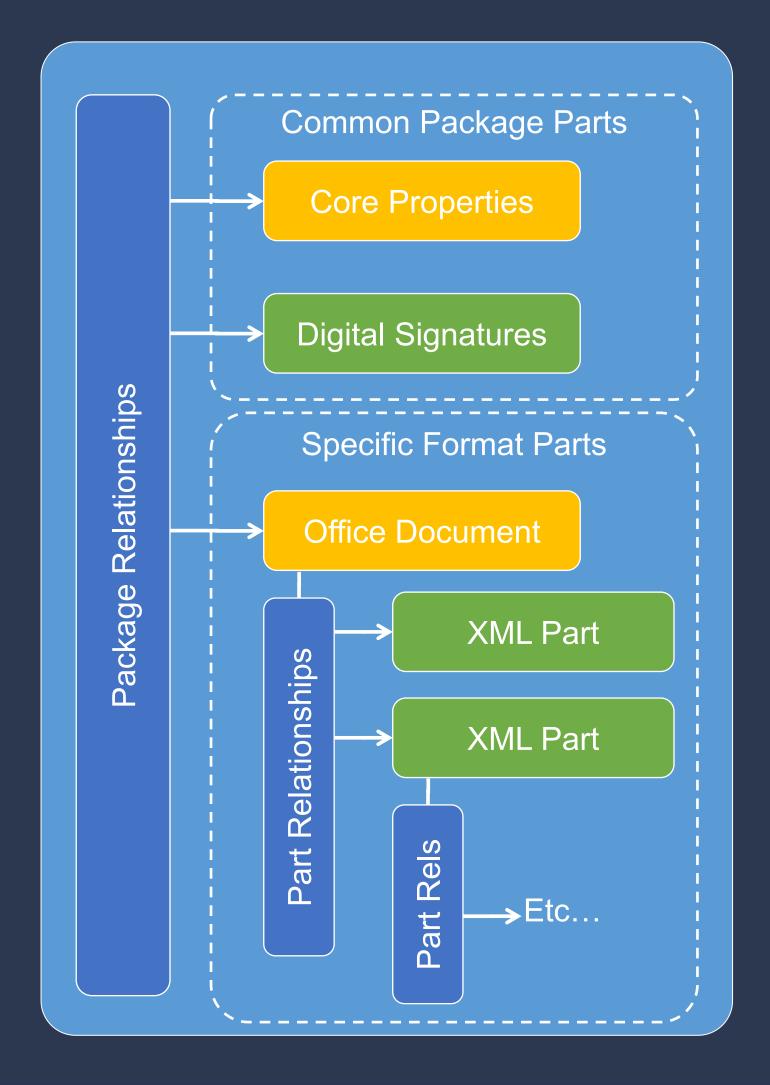


```
// Code generated by xgen. DO NOT EDIT.
package schema
import "encoding/xml"
// MyType1 ...
type MyType1 string
// MyType2 ...
type MyType2 struct {
    XMLName xml.Name xml:"myType2">
    LengthAttr int `xml:"length,attr,omitempty"`
                     `xml:",chardata"`
    Value string
// MyType3 ...
type MyType3 struct {
    XMLName xml.Name `xml:"myType3"`
                      `xml:"length,attr,omitempty"`
    LengthAttr int
                      `xml:",chardata"`
    Value string
// MyType4 ...
type MyType4 struct {
    XMLName xml.Name `xml:"myType4"`
    Title string
                     `xml:"title"`
                     `xml:"blob"`
    Blob string
    Timestamp string `xml:"timestamp"`
// MyType5 ...
type MyType5 string
```





基于文档内存模型的运行时操作



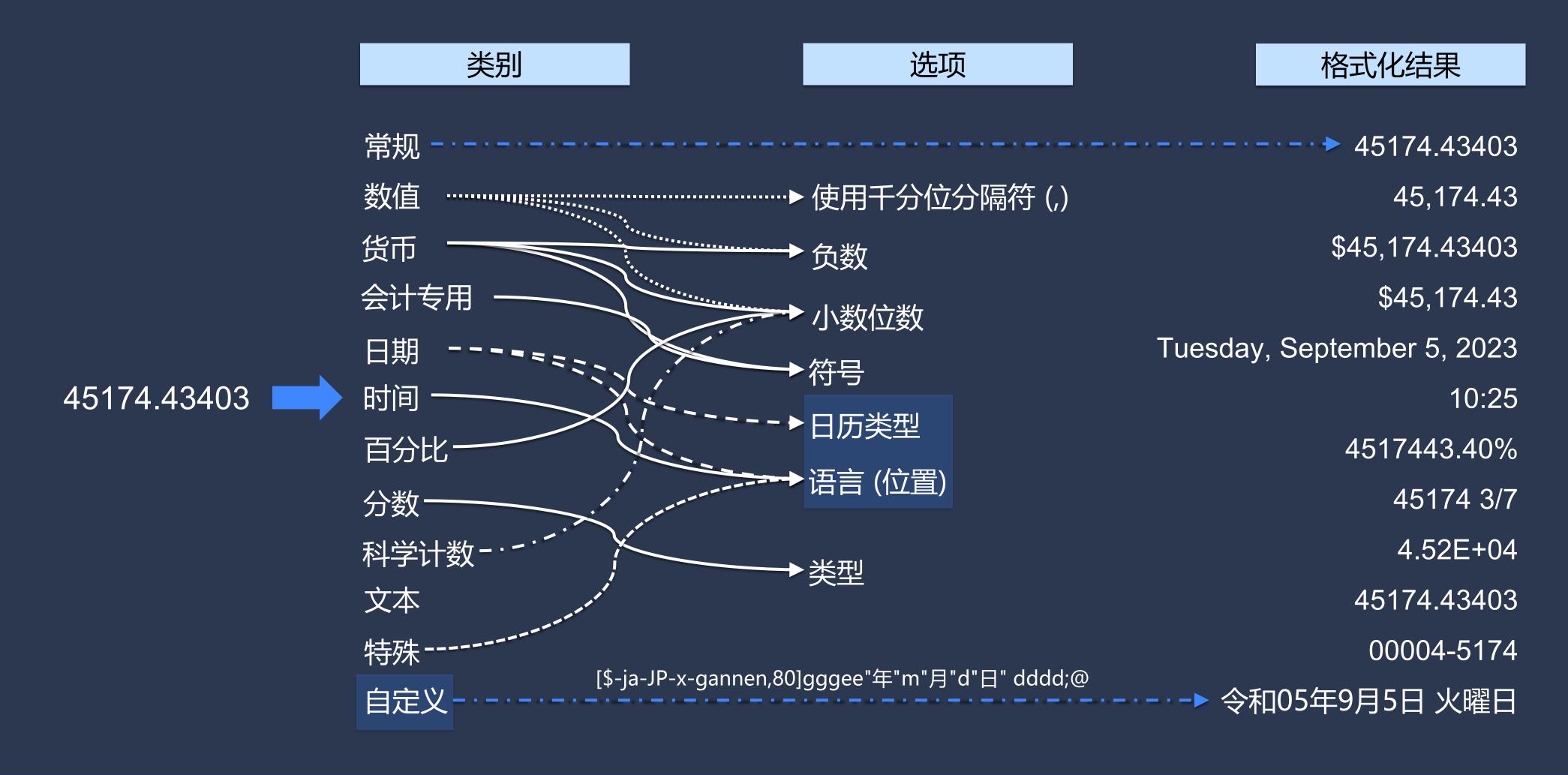
	А	В	С			
1						
2		123				
3						
4						

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<worksheet</pre>
    xmlns="http://schemas.openxmlformats.org/spreadsheetml/2006/main"
    xmlns:mc="http://schemas.openxmlformats.org/markup-compatibility/2006">
    <dimension ref="B2"/>
    <sheetViews>
        <sheetView tabSelected="1" workbookViewId="0" />
    </sheetViews>
    <sheetFormatPr baseColWidth="10" defaultRowHeight="16" />
    <sheetData>
        <row r="2">
            <c r="B2">
                <v>123</v>
            </c>
        </row>
    </sheetData>
   <pageMargins left="0.7" right="0.7" />
</worksheet>
```





电子表格文档中的数字格式







如何应用数字格式?

```
-123.456
    格式化正数的值
                           零的格式
                                  格式化文本内容
              格式化负数的值
#,###.00_);[Red](#.00);0.00;"Sales"@
                        NFP (Number Format Parser)
                        https://github.com/xuri/nfp
ps := nfp.NumberFormatParser()
ps.Parse(`#,###.00_);[Red](#.00);0.00;"Sales"@`)
println(ps.PrettyPrint())
```

```
<Positive>
    # <HashPlaceHolder>
    , <ThousandsSeparator>
    ### <HashPlaceHolder>
    . <DecimalPoint>
    00 <ZeroPlaceHolder>
<Negative>
    Red <Color>
    ( <Literal>
    # <HashPlaceHolder>
    . <DecimalPoint>
    00 <ZeroPlaceHolder>
    ) <Literal>
<Zero>
    0 <ZeroPlaceHolder>
      <DecimalPoint>
    00 <ZeroPlaceHolder>
<Text>
    Sales <Literal>
    @ <TextPlaceHolder>
```





实验评估: 时间与日期类型的数字格式

电子表格文档中的预设数字格式与预设语言地区数字格式 ID 区间范围: 27~36,50~81,自定义数字格式 ID 的起始值为 164

ID	Win Office 07 OS 国家或地区: 美国 OS 区域格式: 中文(简体,中国) Office 编辑语言: 英文(英语,美国)		mac Office 16.3 OS 国家或地区: 美国 OS 区域格式: 美国 Office 编辑语言: 中文(繁体,中国台湾)	mac Office 16.3 OS 国家或地区: 美国 OS 区域格式: 美国 Office 编辑语言: 英文(英语,美国)	Win Office 07 OS 国家或地区: 美国 OS 区域格式: 中文(繁体,中国台湾) Office 编辑语言: 中文(繁体,中国台湾)		Win Office 07 OS 国家或地区: 美国 OS 区域格式: 中文(繁体,中国台湾) Office 编辑语言: 中文(简体,中国)		OS 区域格式: 美国	OS 区域格式: 美国	Win Office 07 OS 国家或地区: 美国 OS 区域格式: 英文(英语,美国) Office 编辑语言: 英文(英语,美国)	Win Office 07 OS 国家或地区: 美国 OS 区域格式: 中文(繁体,中国台湾) Office 编辑语言: 日本语		mac Office 16.3 OS 国家或地区: 美国 OS 区域格式: 美国 Office 编辑语言: 韩国语	Win Office 07 OS 国家或地区: 韩国语 OS 区域格式: 韩国语 Office 编辑语言: 韩国语
LDate	yyyy'年'M'月'd'日'	yyyy'年'M'月'd'日'			yyyy'年'M'月'd'日'	yyyy'年'M'月'd'日'	yyyy'年'MM'月'dd'日'				dddd, MMMM d, yyyy	yyyy'年'M'月'd'日'	yyyy'年'M'月'd'日'		dddd, MMMM d, yyyy
LTime	H:mm:ss	H:mm:ss			tt hh:mm:ss	tt hh:mm:ss	HH:mm:ss				H:mm:ss tt	tt hh:mm:ss	H:mm:ss		H:mm:ss tt
SDate	yyyy/M/d	yyyy/M/d			yyyy/M/d	yyyy/M/d	yyyy-MM-dd				M/d/yyyy	yyyy/M/d	yyyy/MM/dd		M/d/yyyy
STime	H:mm	H:mm			tt hh:mm	tt hh:mm	HH:mm				H:mm tt	tt hh:mm	H:mm		H:mm tt
27	2023/8/24	2023年8月	2023/8/24	4 8/24/23	112/8/24	2023年8月	2023年08月	2023年8月	8.24.R5	8/24/23	8/24/2023	H35.08.24	H35.8.24	8月 24日 2023年	8月 24日 2023年
28	2023/8/24	4 8月24日	2023年8月24日	8/24/23	112年8月24日	8月24日	08月24日	8月24日	8月24日令和5年	8/24/23	8/24/2023	平成35年08月24日	平成35年8月24日	8/24	8/24
29	2023/8/24	8月24日	2023年8月24日	8/24/23	112年8月24日	8月24日	08月24日	8月24日	8月24日令和5年	8/24/23	8/24/2023	平成35年08月24日	平成35年8月24日	8/24	8/24
30	2023/8/24	8/24/23	8/24/23	8/24/23	8/24/23	8/24/23	3 08-24-23	8/24/23	8/24/23	8/24/23	8/24/2023		8/24/23		
31	2023/8/24			8/24/23	2023年8月24日	2023年8月24日	2023年08月24日	1			8/24/2023		2023年8月24日		
32	0:00:00				00時00分	00时00分							0時00分	00시 00분	0시 00분
33	0:00:00					00时00分00利							0時00分00秒	00시 00분 00초	0시 00분 00초
34	0:00:00					上午12时00分				0.0310000000000000000000000000000000000			2023年8月	8-24-23	
35	0:00:00												8月24日		
36	2023/8/24							and the same of th							8月 24日 2023年
50	2023/8/24														
51	2023/8/24					8月24日							平成35年8月24日		
52	2023/8/24					2023年8月							2023年8月		
53	2023/8/24					8月24日							8月24日		
54	2023/8/24				The state of the s	8月24日							平成35年8月24日		
55	2023/8/24					上午12时00分							2023年8月		
56	2023/8/24												8月24日		
57	2023/8/24						to the state of th								
58	2023/8/24												平成35年8月24日		
59	45162														
60	45162.00														
61	45,162														
62	45,162.00														
63	¥45,162 ¥45,162														¥45,162 ¥45.162
64															
65	¥45,162.00 ¥45,162.00														¥45,162.00
66	4516200%														¥45,162.00 4516200%
67	4516200.00%				The state of the s		The second secon						4516200.00%		
68	45162	45162	45162	45162	45162	45162	45162	45162	45162	ଫ୍ଟେଅନ୍ଥର	45162	45162	45162	45162	45162
69	45162	45162	45162	45162	45162	45162	45162	45162	45162	বেও্চচ	45162	45162	45162	45162	45162
70	2023/8/24	The second of th			Secretary program										
71	2023/8/24														
72	24-8月-23														
73 74	24-8月									1000					
75	8月-23				2 12		72.1 20.5								
76	0:00		i de la companya del companya de la companya de la companya del companya de la co						The second secon						
77	0:00:00														
78	2023/8/24 0:00						CONTRACTOR OF THE PROPERTY OF								
79	00:00				722 23								0.000		
80	1083888:00:00												1083888:00:00		
81	00:00.0											5,000 0.000 0.000 0.000 0.000			





数字格式: 国际化 & 本地化







Excelize 对时间与日期类型数字格式的解决方案

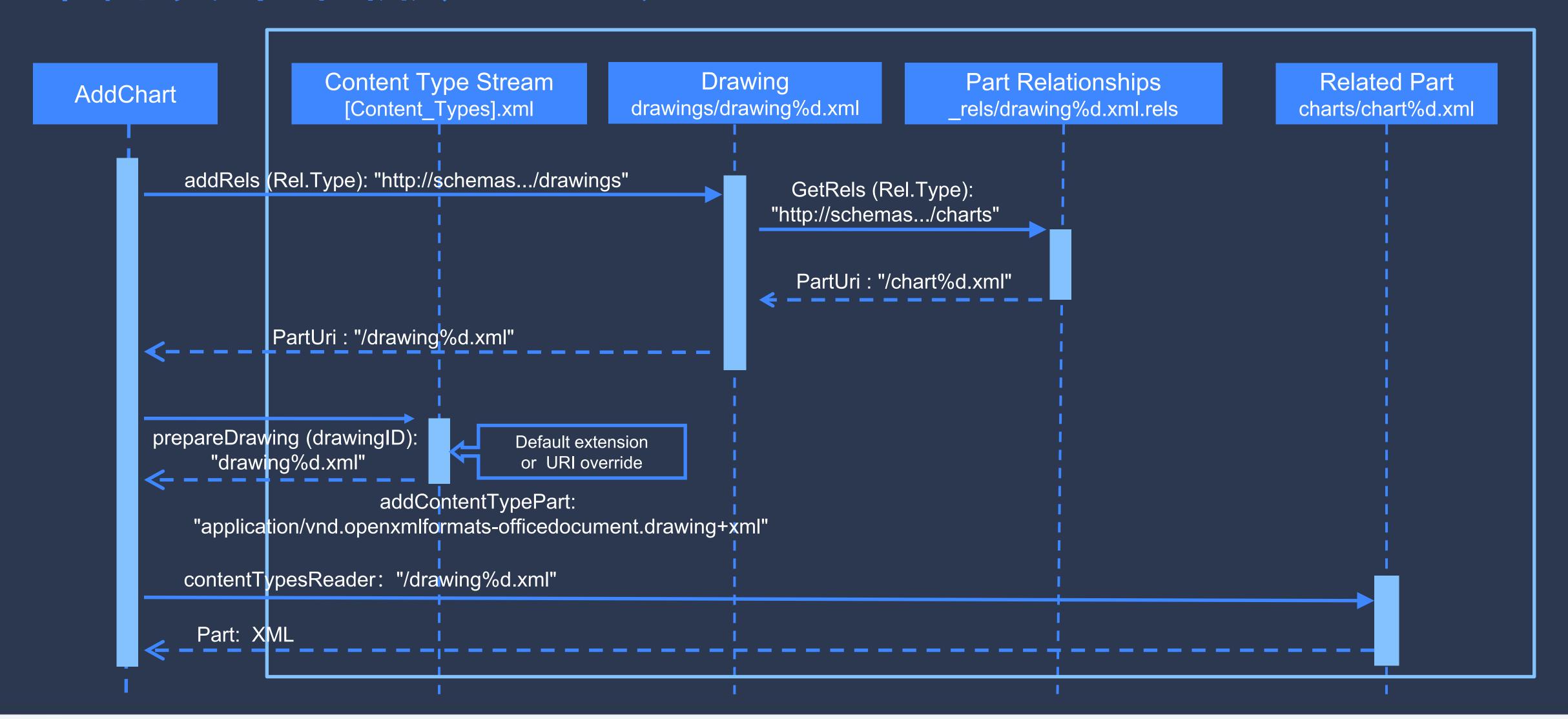
参考标准 [MS-LCID]: Windows Language Code Identifier Reference Excelize 已覆盖包含超过 810 个语言地区标识的时间与日期类型数字格式

```
时间和语言 > 语言和区域 > 区域格式
f, err := excelize.OpenFile("Book1.xlsx",
    excelize.Options{
                                                                          一周的第一天
                             excelize.CultureNameZhCN,
         CultureInfo:
                                                                                                           2017/4/5
         ShortDatePattern: "dd/mm/yy",
                                                                                                           2017年4月5日 ~
         LongDatePattern: "yyyy-mm-dd",
         LongTimePattern: "hh:mm:ss", -
                                                                                                           9:40:07
    },
                                                                          标准数字
                                                                                                           0123456789 ~
                                                                          排序方法
                                                                         ₽ 提供反馈
```





架构设计:图形处理系统

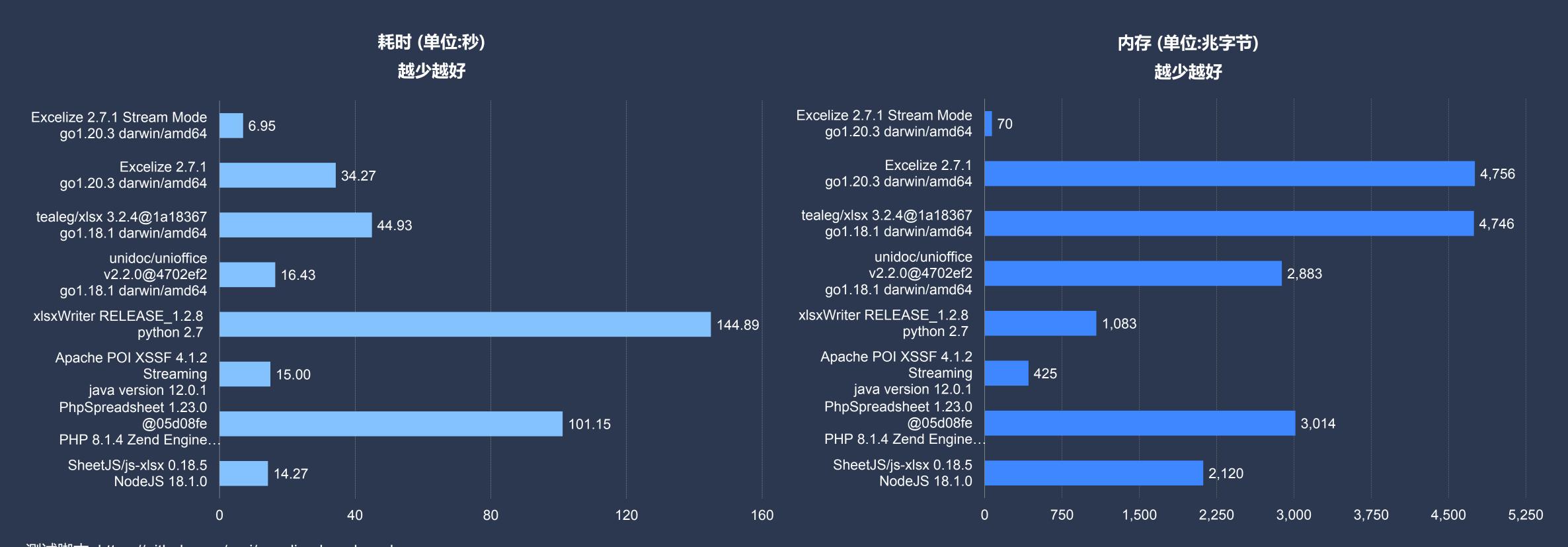






性能表现

下图展示了 Go, Python, Java, PHP 和 NodeJS 语言中典型 Excel 开源基础库,基于普通个人计算机 (2.6 GHz 6-Core Intel Core i7, 16 GB 2667 MHz DDR4, 500GB SSD, macOS Monterey 12.3.1) 生成 50 列 102400 行纯文本单元格的性能表现



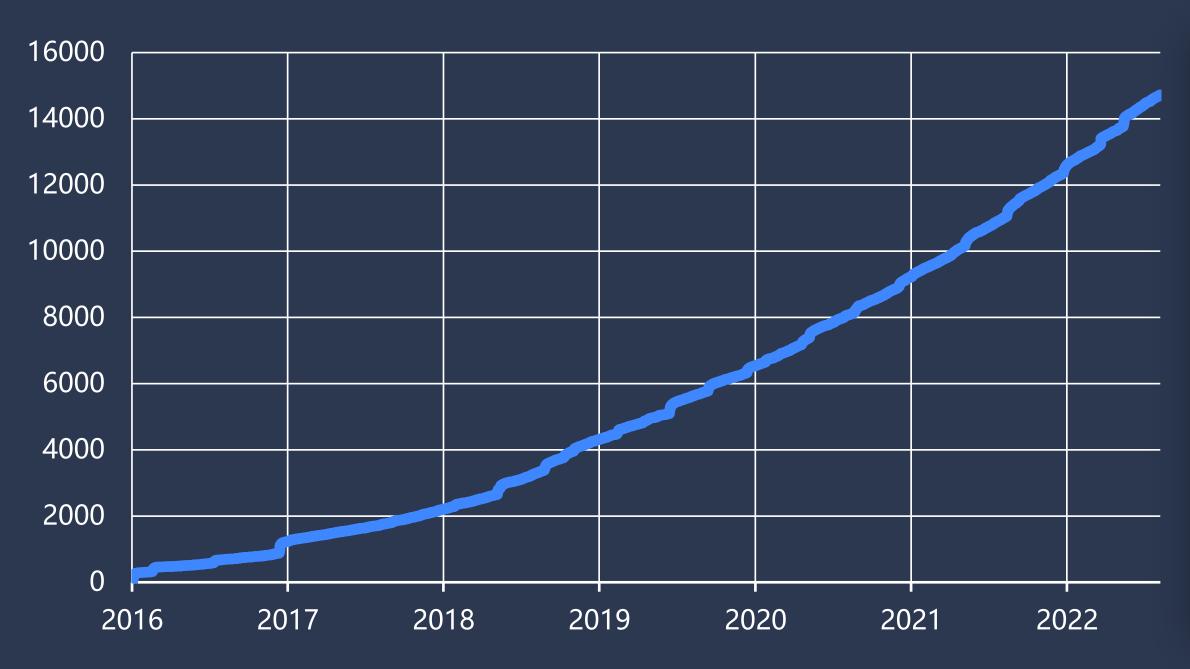
测试脚本: https://github.com/xuri/excelize-benchmark 函数性能: https://xuri.me/excelize/en/performance.html





开源数据洞察

Excelize GitHub Star 趋势



Excelize 全球范围内贡献者分布



* 截至 2023 年 9 月,GitHub Star 1.5 万,有超过 50 个国家和地区的 400 多名贡献者,其中有 170 余人参与了代码贡献





WebAssembly 跨语言调用

npm install --save excelize-wasm

```
const { init } = require('excelize-wasm');
const fs = require('fs');
init('./node_modules/excelize-wasm/excelize.wasm.gz').then((excelize) => {
    const f = excelize.NewFile();
    const { index } = f.NewSheet('Sheet2');
   f.SetCellValue('Sheet2', 'A2', 'Hello world.'); // 单元格赋值
   f.SetCellValue('Sheet1', 'B2', 100);
   f.SetActiveSheet(index); // 设置活动工作表
    const { buffer, error } = f.WriteToBuffer();
   if (error) {
       console.log(error);
       return;
    fs.writeFile('Book1.xlsx', buffer, 'binary', (error) => {
       if (error) {
            console.log(error);
});
```





总结展望

功能规划

更多接口将提供并发安全支持

对包含大规模数据文档读写性能的持续优化

公式计算引擎增加更多计算函数

复杂条件格式的设置、数据预测支持

灵活控制图表样式

数据透视图、切片器功能

更加丰富的数字格式表达式解析

提供更多可用流式读写功能





THANKS

软件正在重新定义世界 Software Is Redefining The World



Excelize 续日



GitHub 代码库



