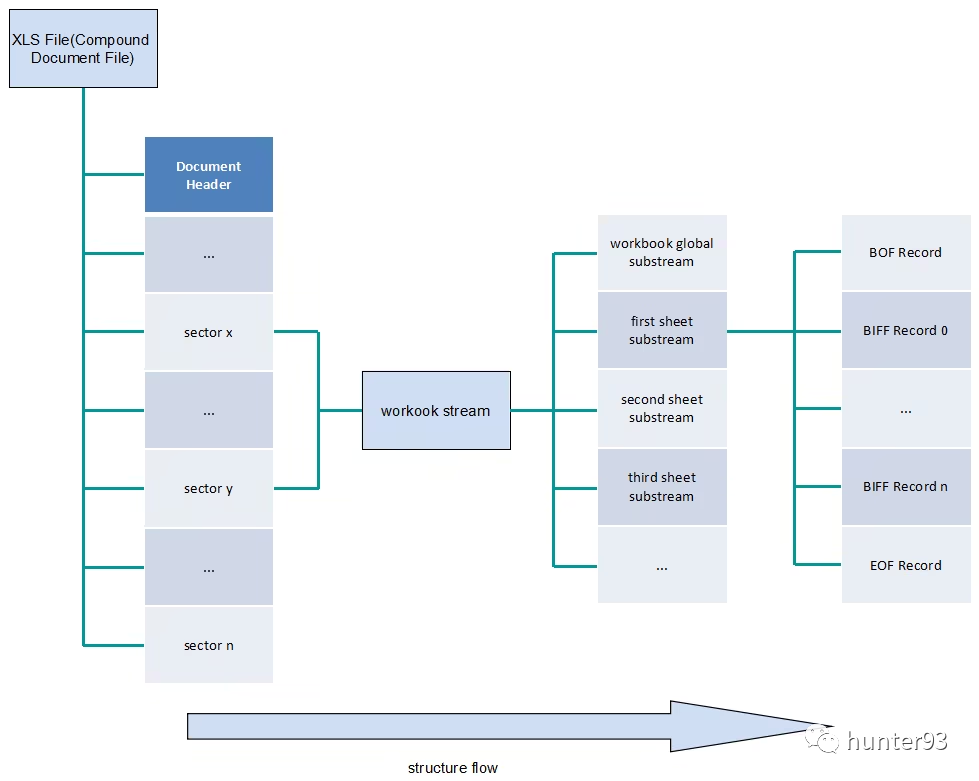
**文件格式:**

结构组成: 文件头 + 多个扇区 + 数据流



****文件头（512字节）****

包含魔数D0 CF 11 E0 A1 B1 1A E1、扇区大小（通常512字节）、FAT表位置等关键元数据

****扇区组织****

数据通过****扇区链****存储，由****扇区分配表（FAT）​****管理链式结构，FAT本身也存储在扇区中

****目录流****

红黑树结构存储所有条目（128字节/条目），包含名称、类型（流/存储）、起始扇区ID、流大小等信息

****短流与长流****​≤4096字节的流使用****短流存储****，通过短扇区分配表（MiniFAT）管理.

libxls

**项目架构:**

libxls/ # 项目根

├── include/ # 公共头文件

│ ├── libxls/ # 对外 C API（xls.h、brdb.h…）

│ └── xls2csv.h # CLI 工具专用

├── src/ # 库实现

│ ├── ole.c/.h # OLE2 容器解析

│ ├── xls.c/.h # WorkBook/WorkSheet 高层封装

│ ├── biff.c/.h # BIFF 记录级解析

│ ├── sst.c/.h # Shared String Table

│ ├── utf16.c/.h # 字符集转换

│ └── endian.c/.h # 大小端适配

├── test/ # 单元测试

├── fuzz/ # libFuzzer 语料

├── xls2csv.c # CLI 主程序

├── bootstrap # autotools 生成脚本

└── configure.ac # autotools 配置

**工作流程:**

[CLI] xls2csv.c

│ xls\_open\_file()

▼

[src/xls.c] 高层 API

│ 1. 委托 ole.c 打开 OLE2 容器

│ 2. 读取 Workbook Global Substream

│ 3. 解析 SST → 缓存所有共享字符串

▼

[src/ole.c] 容器层

│ 读取目录/扇区链

▼

[src/biff.c] 记录层

│ 按 BIFF8 格式解析 BOF/ROW/CELL/FORMULA…

▼

[src/utf16.c/endian.c]

│ 字节序 & 编码转换

▼

回到 [xls.c] 组装成 xlsWorkSheet

参考文献:

1. <https://mp.weixin.qq.com/s?__biz=MzI0MDU4NjU2OQ==&mid=2247483697&idx=1&sn=417e4b0c5e9cd5f0ba8cd6f01541152a&chksm=e88464955a7c9da268c25e1a3d79e36d5379c68cfe8acd2ccf19e538657368eaf2cf3ce810f4#rd>
2. <https://github.com/libxls/libxls>
3. <https://learn.microsoft.com/en-us/openspecs/office_file_formats/ms-xls/cd03cb5f-ca02-4934-a391-bb674cb8aa06>