

# 数据库设计说明书

## 一、引言

### 1. 编写目的

统一 Recubator 项目数据库的设计标准与实现规范，确保数据结构的完整性、一致性与可维护性，为前后端开发、测试、运维提供明确依据，支撑系统本地存储、高效查询及后续功能扩展，减少因设计偏差导致的返工。

### 2. 项目背景

本数据库为 Windows 桌面端软件 Recubator 专属，用于存储用户记录、媒体文件、标签、摘要、网页快照等核心业务数据。软件聚焦本地闭环使用，仅 LLM 相关功能需联网调用 API，数据库采用嵌入式部署模式，无需独立服务器，旨在解决用户本地笔记数据高效管理、关联查询及安全存储的需求。

### 3. 参考资料

- 项目用例规约
- Recubator 表结构设计文档
- 软件功能模块概述
- SQLite/H2 官方技术手册

## 二、需求分析

### 1. 业务需求分析

核心业务围绕用户本地笔记管理展开，关键流程包括：

- 记录创建：支持文本输入、多类型文件（图片、视频、文档等）拖拽上传，系统自动保存草稿。
- 标签管理：支持系统自动推荐标签、用户手动创建标签，记录与标签多对多关联。
- 内容解析：对含网页 / 音视频的记录，自动提取元数据、生成摘要与网页快照。
- 数据关联：实现记录与媒体、记录与标签、记录与摘要 / 快照的关联查询。
- 重复检测：新增记录时计算与现有记录的相似度， $\geq 85\%$  触发重复提醒。

### 2. 用户需求分析

目标用户为个人桌面端笔记使用者，核心数据操作需求：

- 数据读写：快速创建、编辑、查询、删除个人记录及关联数据。
- 标签操作：为记录添加 / 删除标签，通过标签筛选记录。
- 内容查看：查看记录关联的媒体文件、摘要及网页快照。
- 数据安全：本地存储保障数据隐私，支持数据库备份与恢复。

### 3. 非功能需求分析

- 性能：单用户场景下，查询响应时间≤300ms，记录创建 / 编辑操作响应时间≤500ms。
- 可靠性：数据存储无丢失，支持自动备份，故障恢复后数据完整性不受影响。
- 安全性：数据库文件加密存储，防止未授权访问。
- 扩展性：支持每年 10% 数据量增长，表结构变更不影响现有业务功能。
- 兼容性：适配 SQLite 与 H2 嵌入式数据库，满足不同用户性能需求。

## 三、外部设计

### 1. 数据库标识体系

- 唯一编码：Recubator\_DB
- 适用范围：支撑 Recubator 记录管理、标签管理、智能解析等核心模块

### 2. 关联应用程序清单

应用程序名称	调用方式	核心操作
Recubator 前端 (Electron)	JDBC	记录、标签、媒体等数据的增删改查
本地数据备份工具	命令行	数据库全量备份、增量备份与恢复

### 3. 设计约定

- 命名规则：表名采用“功能模块\_实体名称”（如 record、media、tag），字段名用小写 + 下划线命名（如 record\_id、create\_time），避免数据库保留字。
- 格式标准：时间字段统一使用 DATETIME 类型（格式：YYYY-MM-DD HH:MM:SS），编码统一为 UTF-8，空值统一用 NULL 表示。
- 数据约束：数值型字段明确取值范围，字符串字段限定长度，非空字段强制校验。

### 4. 支持软件清单

类别	软件名称及版本	用途
DBMS	SQLite 3.40+ / H2 2.2+	嵌入式数据库核心存储
开发依赖	MyBatis-Plus 3.5+	后端数据访问层框架

## 四、概念结构设计

### 1. 实体识别与属性定义

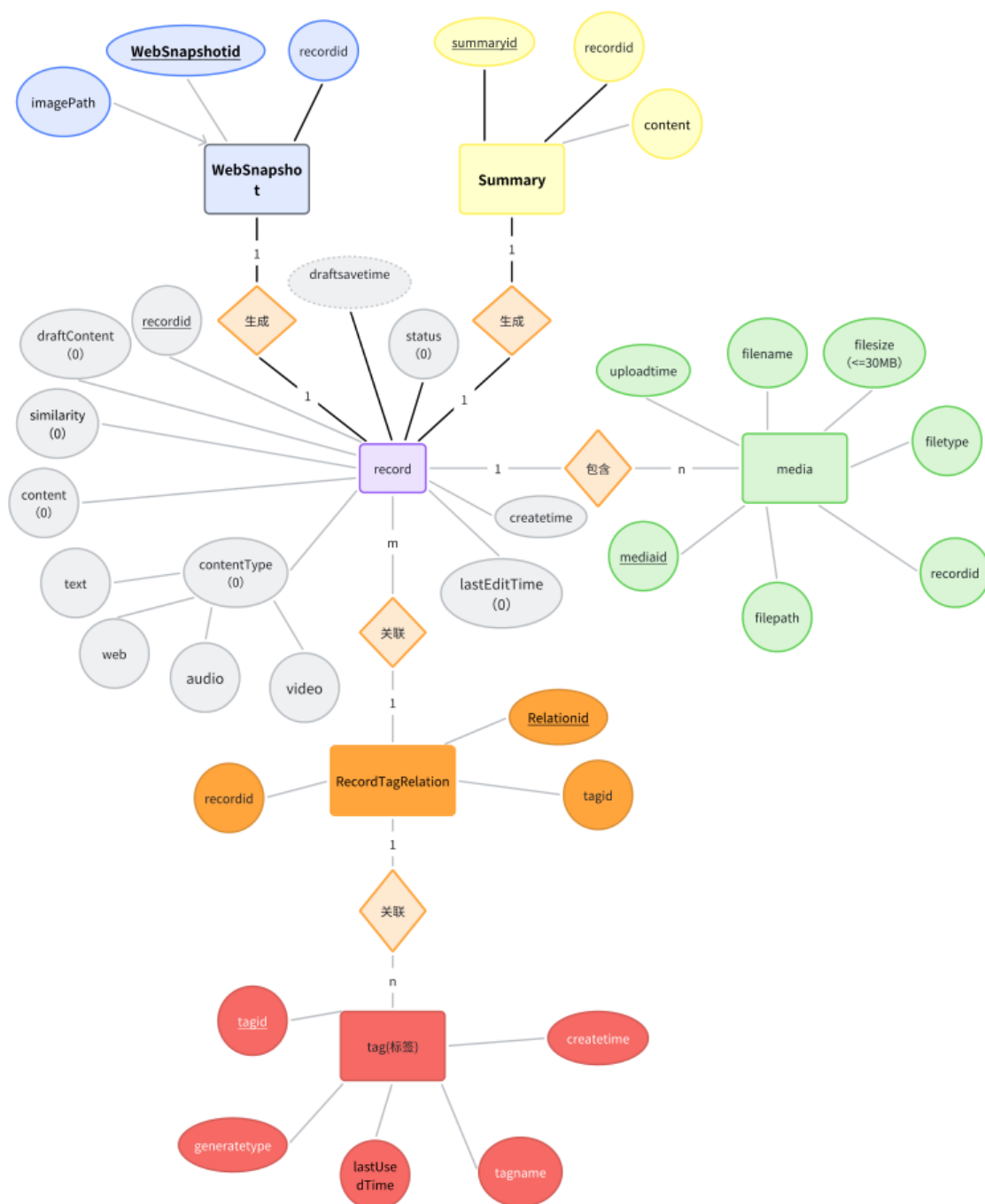
- 记录 (Record)：核心实体，属性包括 recordId、content、draftContent、contentType、status、similarity、createTime、lastEditTime、draftSaveTime。
- 媒体 (Media)：关联实体，属性包括 mediaId、recordId、fileName、filePath、fileType、fileSize、uploadTime。
- 标签 (Tag)：核心实体，属性包括 tagId、tagName、generateType、createTime、lastUsedTime。

- 记录标签关联 (RecordTagRelation)：关联实体，属性包括 relationId、recordId、tagId。
- 摘要 (Summary)：关联实体，属性包括 summaryId、recordId、content、metadata、generateTime。
- 网页快照 (WebSnapshot)：关联实体，属性包括 snapshotId、recordId、imagePath、generateTime。

## 2. 实体关系定义

- 记录与媒体：一对多关系（1 条记录可关联多个媒体文件，1 个媒体文件仅关联 1 条记录）。
- 记录与标签：多对多关系（1 条记录可关联多个标签，1 个标签可关联多条记录），通过记录标签关联表实现。
- 记录与摘要：一对一关系（1 条记录最多对应 1 条摘要，1 条摘要仅关联 1 条记录）。
- 记录与网页快照：一对一关系（1 条含网页的记录对应 1 个快照，1 个快照仅关联 1 条记录）。

## 3. E-R 图



- Record 与 Media：包含（1:N）
- Record 与 Tag：关联（M:N），通过 RecordTagRelation 连接
- Record 与 Summary：生成（1:1）
- Record 与 WebSnapshot：生成（1:1）

## 五、逻辑结构设计

### 1. 表结构详细设计

#### (1) 记录表（Record）

字段名	数据类型	主键	外键	约束条件	业务说明
recordId	INT	√	-	自增、非空、唯一	记录唯一标识
content	TEXT	-	-	可空	记录文本内容，允许纯媒体记录
draftContent	TEXT	-	-	可空	草稿内容，系统每 5 秒自动保存
contentType	JSON	-	-	可空	标识记录包含的内容类型（文本 / 网页 / 视频等）
status	VARCHAR(20)	-	-	非空，枚举（draft/saved/edited）	记录状态
similarity	DECIMAL(5,2)	-	-	可空，默认 0.00	与现有记录的相似度（%），≥85% 触发重复提醒
createTime	DATETIME	-	-	非空	记录创建时间
lastEditTime	DATETIME	-	-	可空	最后编辑时间
draftSaveTime	DATETIME	-	-	可空	草稿最后保存时间（派生属性）

#### (2) 媒体表（Media）

字段名	数据类型	主键	外键	约束条件	业务说明
mediaId	INT	√	-	自增、非空、唯一	媒体唯一标识
recordId	INT	-	√ (关联 Record.recordId)	非空	关联的记录 ID

字段名	数据类型	主键	外键	约束条件	业务说明
fileName	VARCHAR(200)	-	-	非空	媒体文件名 (如“笔记图片.png”)
filePath	VARCHAR(500)	-	-	非空	媒体本地存储路径
fileType	VARCHAR(50)	-	-	非空，枚举 (text/link/image/video/audio/document/web)	媒体类型
fileSize	DECIMAL(10,2)	-	-	非空，≤30.00	媒体大小 (单位：MB)，不超过 30MB
uploadTime	DATETIME	-	-	非空	媒体上传时间

(3) 标签表 (Tag)

字段名	数据类型	主键	外键	约束条件	业务说明
tagId	INT	√	-	自增、非空、唯一	标签唯一标识
tagName	VARCHAR(50)	-	-	非空、唯一	标签名称，不重复创建
generateType	VARCHAR(20)	-	-	非空，枚举 (auto/manual)	生成类型（系统自动推荐 / 用户手动创建）
createTime	DATETIME	-	-	非空	标签创建时间
lastUsedTime	DATETIME	-	-	可空	标签最近使用时间，用于候选标签展示

(4) 记录标签关联表 (RecordTagRelation)

字段名	数据类型	主键	外键	约束条件	业务说明
relationId	INT	√	-	自增、非空、唯一	关联唯一标识
recordId	INT	-	√ (关联 Record.recordId)	非空	关联的记录 ID
tagId	INT	-	√ (关联 Tag.tagId)	非空	关联的标签 ID
-	-	-	-	联合唯一索引 (recordId, tagId)	确保一条记录不重复关联同一标签

(5) 摘要表 (Summary)

字段名	数据类型	主键	外键	约束条件	业务说明
summaryId	INT	√	-	自增、非空、唯一	摘要唯一标识
recordId	INT	-	√ (关联 Record.recordId)	非空、唯一	关联的记录 ID (1:1 关系)
content	TEXT	-	-	非空	自动生成的摘要内容
metadata	JSON	-	-	非空	存储内容元数据 (标题、封面、时长等)
generateTime	DATETIME	-	-	非空	摘要生成时间

(6) 网页快照表 (WebSnapshot)

字段名	数据类型	主键	外键	约束条件	业务说明
snapshotId	INT	√	-	自增、非空、唯一	快照唯一标识
recordId	INT	-	√ (关联 Record.recordId)	非空、唯一	关联的记录 ID (1:1 关系)
imagePath	VARCHAR(500)	-	-	非空	网页快照图片本地存储路径
generateTime	DATETIME	-	-	非空	快照生成时间 (仅记录含网页时生成)

2. 表间关系与约束

- 外键约束：所有外键字段均关联对应主表主键，删除主表记录时，级联删除关联表相关记录（如删除 Record 时，级联删除 Media、Summary、WebSnapshot 及 RecordTagRelation 中对应记录）。
- 唯一性约束：关键业务字段（如 tagName、recordId 与 tagId 组合）设置唯一约束，避免重复数据。
- 检查约束：fileSize 字段设置≤30.00 的检查约束，确保文件大小符合业务规则。

### 3. 规范化处理

所有表结构均遵循第三范式（3NF），消除数据冗余与更新异常：

- 每个表仅存储单一实体的相关数据（如 Record 表仅存储记录核心信息，媒体信息单独存储在 Media 表）。
- 非主键字段仅依赖于主键，不依赖其他非主键字段（如 draftSaveTime 仅依赖 recordId，不依赖其他字段）。
- 多对多关系通过中间表（RecordTagRelation）拆分，避免直接关联导致的数据冗余。

## 六、物理结构设计

### 1. DBMS 选型与配置

#### (1) 默认选型：SQLite 3.40+

- 选型理由：轻量级嵌入式数据库，单文件存储，无需独立服务，适配桌面端本地部署需求，占用资源少。

#### (2) 可选选型：H2 2.2+

- 选型理由：支持文件模式与内存模式，性能优于 SQLite，支持更多 SQL 特性，适合数据量较大（10GB-50GB）的用户。

### 2. 存储结构设计

- 存储路径：默认存储于 `C:\Users\{用户名}\AppData\Roaming\Recubator\db\recubator.db` (SQLite)，用户可手动修改存储路径。

### 3. 索引设计

表名	索引字段	索引类型	用途
Record	createTime	普通索引	按创建时间排序查询记录
Record	status	普通索引	按状态筛选记录（如查询所有草稿）
Media	recordId	普通索引	快速查询某条记录的关联媒体
Tag	lastUsedTime	普通索引	快速筛选最近使用的标签
RecordTagRelation	(recordId, tagId)	联合唯一索引	确保记录与标签关联唯一，提升关联查询效率
Summary	recordId	唯一索引	快速查询某条记录的摘要，确保 1:1 关系
WebSnapshot	recordId	唯一索引	快速查询某条记录的快照，确保 1:1 关系

# 七、运用设计

## 1. 数据字典设计

数据项名称	英文标识	数据类型	取值范围	业务规则	更新频率
记录ID	recordId	INT	1~2^31-1	自增, 唯一标识一条记录	每次创建记录时更新
记录内容	content	TEXT	无长度限制	可选, 允许纯媒体记录	创建 / 编辑记录时更新
记录状态	status	VARCHAR(20)	draft/saved/edited	非空, 标识记录当前状态	创建 / 编辑 / 保存记录时更新
媒体ID	mediaId	INT	1~2^31-1	自增, 唯一标识一个媒体文件	上传媒体时更新



数据项名称	英文标识	数据类型	取值范围	业务规则	更新频率
媒体类型	fileType	VARCHAR(50)	text/link/image/video/audio/document/web	非空，标识媒体文件类型	上传媒体时更新
标签ID	tagId	INT	1~2^31-1	自增，唯一标识一个标签	创建标签时更新
标签名称	tagName	VARCHAR(50)	非空，无特殊字符	唯一，不重复创建	创建/修改标签时更新
摘要ID	summaryId	INT	1~2^31-1	自增，唯一标识一条摘要	生成摘要时更新
快照ID	snapshotId	INT	1~2^31-1	自增，唯一标识一个快照	生成快照时更新

## 2. 安全保密设计

- 访问控制：基于用户操作系统权限控制数据库文件访问，仅软件运行账户可读写数据库文件。
- 数据加密：数据库文件采用 SQLCipher (SQLite) /AES-256 (H2) 加密存储，加密密钥通过软件配置文件加密保存。

## 八、维护

### 日常维护计划

维护任务	周期	操作内容	执行
索引优化	每周	重建碎片化率≥30% 的索引，提升查询效率	软件自动执行
日志清理	每月	删除超过 30 天的审计日志，释放磁盘空间	软件自动执行
数据库压缩	每季度	压缩数据库文件，清除冗余空间	软件自动执行
完整性校验	每月	执行数据库完整性校验，发现问题自动触发备份恢复	软件自动执行

## 附录

### 1. 修订历史

修订版本	修订时间	修订人	修订内容
V1.0	2025-11-19	Justin, JZK	初始版本，完成核心章节编写