

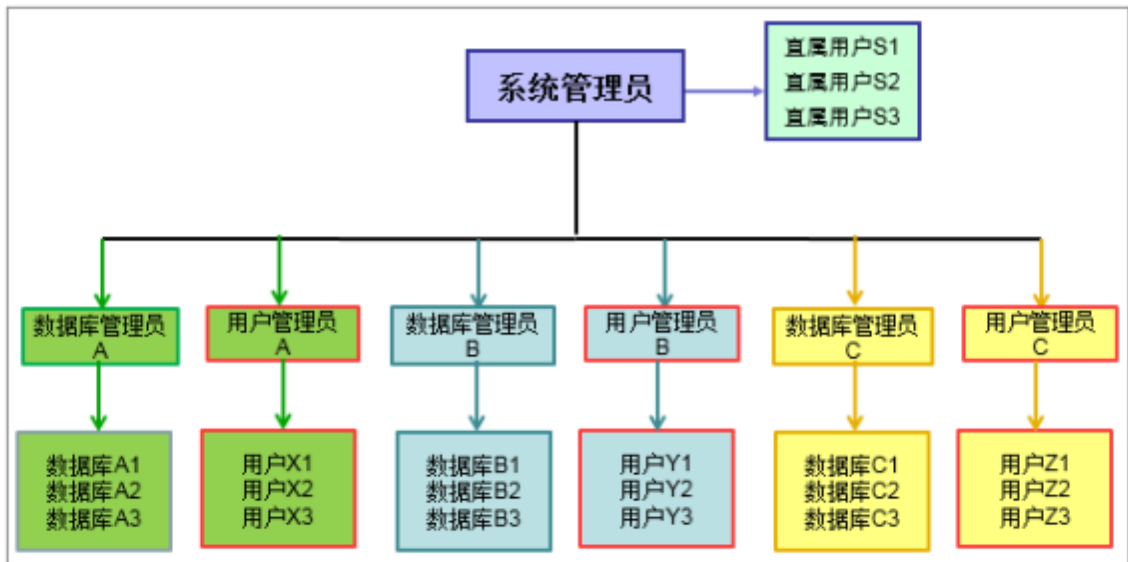


图 4.1 用户管理系统示意图

4.2 用户组管理系统

用户组用于数据库访问授权管理，其作用是在访问数据库时不同用户组担当不同的角色。编组方式设计巧妙，适用于严密复杂的数据库访问授权，其主要特点是：

- (1) 系统管理员、用户管理员创建和管理用户组；
- (2) 系统管理员创建和管理的用户组数量无限制；
- (3) 每个用户管理员创建和管理的用户组数量无限制；
- (4) 用户组成员配置方便灵活。用户组成员可以是系统管理员、数据库管理员、用户管理员和一般用户；
- (5) 系统管理员、同一个数据库管理员、同一个用户管理员和同一个一般用户可以作为多个用户组的成员，也就是说，访问同一个数据库或不同的数据库时可以担当不同的角色。
- (6) 用户所在的用户组被授权访问数据库的情况下，就不必对用户组的成员访问数据库再进行授权。

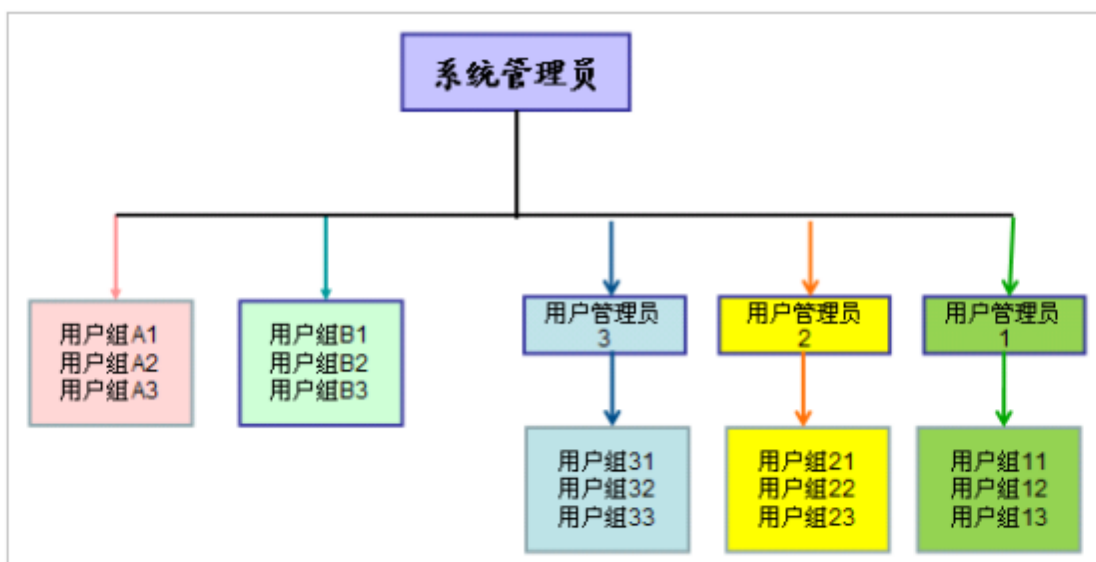


图 4.2 用户组管理系统示意图

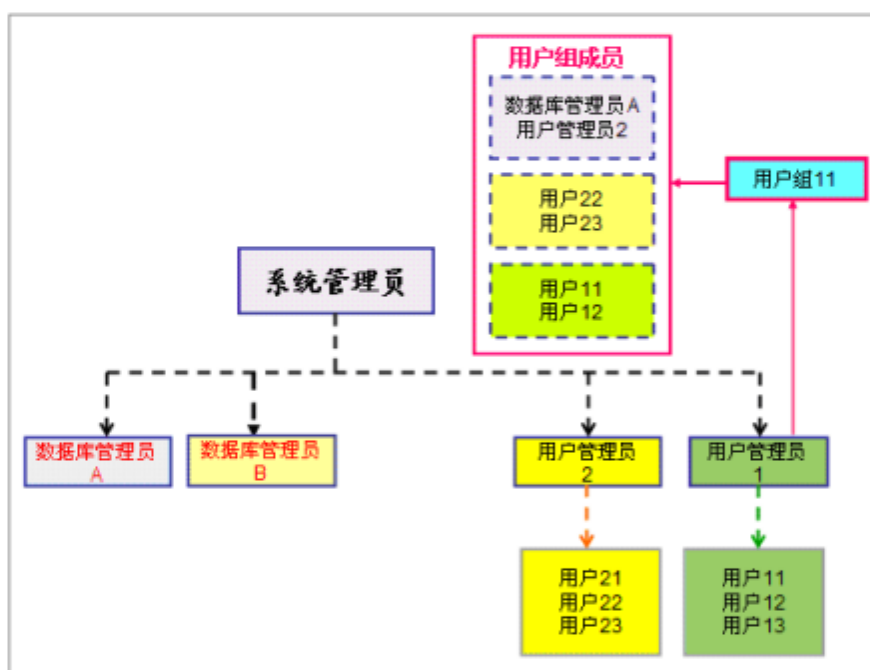


图 4.3 用户组成员构成示意图

4.3 数据库管理系统

4.3.1 管理数据库的用户

管理数据库的用户有两类：系统管理员、数据库管理员。

- (1) 系统管理员创建和管理数据库，但无权管理数据库管理员管理的数据库；
- (2) 数据库管理员创建和管理数据库，但无权管理系统管理员管理的数据库；
- (3) 数据库管理员的数量无限制；
- (4) 系统管理员创建和管理的数据库数量无限制；

(5) 数据库管理员创建和管理的数据库数量无限制；

(6) 其他用户访问数据库，需要管理该数据库的系统管理员或数据库管理员的授权。

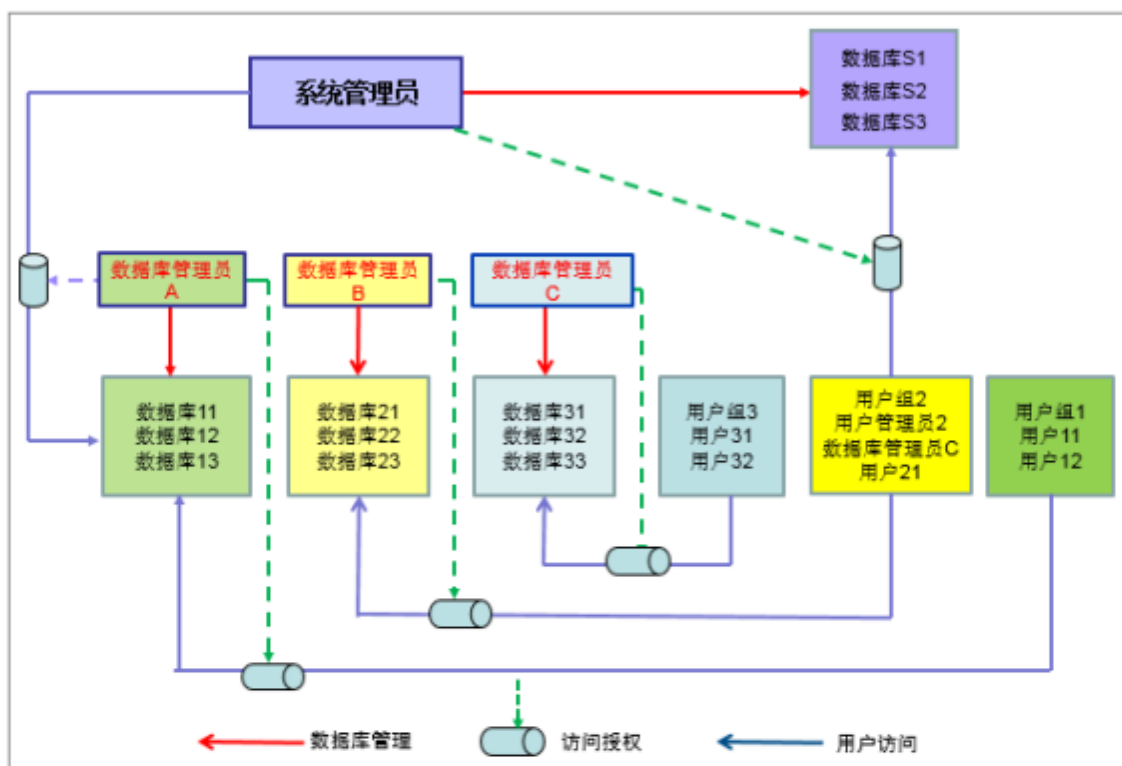


图 4.4 数据库管理示意图

4.3.2 数据库管理用户界面

数据库管理用户界面有 9 项：新建数据库、导入定义文件、导出定义文件、复制数据库、修改数据库、删除数据库、字段编辑、访问权限、转移权限。



4.3.3 新建数据库

新建数据库包括：数据库名、数据库描述、唯一属性字段、可检索字符，选择数据库基本文件、索引文件和词汇索引文件的存储路径以及日志文件的存储路径，选择中文分词类型。

新建数据库对话框的截图，显示了以下配置：

- 名称: file1
- 描述: 文件1
- 唯一属性字段: title
- 基本文件: D://database/file1.baf (浏览按钮)
- 基本索引文件: E://database/file1.bif (浏览按钮)
- 词汇索引文件: E://database/file1.vif (浏览按钮)
- 日志文件: F://database/file1.log (浏览按钮)
- 可检索字符: @-
- 中文分词类型: 全部匹配 (下拉菜单)
- 底部按钮: 保存, 取消

图 4.5 新建数据库对话框

说明：

(1) 数据库名由英文字母、数字和下划线组成，须由英文字母开头，最长 16 个字符。

(2) 描述使用的字符不限，可使用中文。

(3) 可检索字符是一些字符经过定义可被检索，例如：- / * @ ... 。

(4) 中文自动分词选择：单汉字 (N)、全部匹配 (A)、最大匹配 (M)、多字词 (W) 等四种分词方法已经达到实用的水平。以“中华人民共和国”为例说明这四种中文分词方法：

单汉字：切出 7 个单字词：/中/华/人/民/共/和/国/

全部匹配：切出 14 个词：/中华人民共和国/人民共和国/人民/共和国/共和/中华/华人/中/华/人/民/共/和/国/

最大匹配：切出 1 个词：/中华人民共和国/

多字词：切出 7 个词：中华人民共和国/中华/人民共和国/人民/共和国/共和/国

4.3.4 导入定义文件

按照数据库的要求和字段要求编写数据库结构定义文件，导入定义文件用于新建数据库。数据库结构定义文件包括：数据库名、数据库描述、唯一属性字段、

可检索字符、中文分词类型、字段号、字段名、字段名描述、字段类型、子字段、索引类型。

数据库基本文件、索引文件和词汇索引文件的存储路径以及日志文件的存储路径默认为 C:\Windows\system32，可根据具体要求修改存储路径。

4.3.5 导出定义文件

导出定义文件包括：数据库名、数据库描述、唯一属性字段、可检索字符、中文分词类型、字段号、字段名、字段名描述、字段类型、子字段、索引类型。

4.3.6 复制数据库

复制数据库是复制数据库的结构。包括：唯一属性字段、可检索字符、数据库基本文件、索引文件和词汇索引文件的存储路径以及日志文件的存储路径，中文分词类型、字段号、字段名、字段名描述、字段类型、子字段、索引类型。

4.3.7 修改数据库

修改数据库规定：

(1) 数据库名、唯一属性字段不可以修改。

(2) 数据库描述可以修改。

(3) 可检索字符、中文分词类型可以修改，修改后需重新索引。

(4) 数据库基本文件、索引文件和词汇索引文件的存储路径以及日志文件的存储路径可以修改，修改后需将数据库基本文件、索引文件和词汇索引文件以及日志文件移动到新的路径。

(5) 字段号、字段名、字段类型不可修改。

(6) 字段名描述可以修改。

(7) 子字段、索引类型可以修改，修改后需重新索引。

4.3.8 删除

数据库无数据，可被删除。如果数据库有数据，则不能被删除。

4.3.9 字段编辑

数据库字段编辑用户界面有 3 项：新建、修改、删除。



4.3.9.1 新建字段

新建字段包括：字段名、字段描述，选择子字段限制、子记录字段、字段类型、

是否索引，见图 4.6。

The dialog box titled "新建字段" (New Field) contains the following fields and options:

- 名称 (Name): A text input field with a red warning icon.
- 描述 (Description): A text input field.
- 子字段 (Sub-fields): Two input fields, the first containing "0" and the second containing "至0".
- 子记录 (Sub-records): A checkbox.
- 类型 (Type): A dropdown menu with "短语" (Phrase) selected.
- 索引 (Index): A checkbox.
- 词组索引 (Phrase Index): A checkbox.
- 字段特定索引 (Field-specific Index): A checkbox.
- 文本索引 (Text Index): A checkbox.
- Buttons: "保存" (Save) and "取消" (Cancel) at the bottom right.

图 4.6 新建字段对话框

说明：

(1) 字段名由英文字母和下划线组成，须由英文字母开头，最长 16 个西文字符。

(2) 描述使用的字符不限，可使用中文。

(3) 子字段限制

空白 无限制

0..4* 允许没有内容，但最多四个

1..1* 内容不能缺，最多 1 个

1..* 此字段必须有一个内容

1..2* 此字段必须有一个，多至 2 个子字段内容

(4) 字段类型有 7 种：短语 (Phrase)、整数 (Integer)、数值 (Number)、文本 (Text)、日期 (Date)、时间 (Time)、字符串 (String)。

短语 (Phrase)：如名称、标题、地址、关键词、电话

整数 (Integer)：-2,147,483,647 至 +2,147,483,647

数值 (Number)：整数、实数 -1.7E+37 至 1.7E+37

文本 (Text)：按段落、句子、词格式编写的自由文本

日期 (Date): 按 YYYY-MM-DD, YYMM.DD 或 YY-MM

时间 (Time): 按 HH:MM:SS, HH-MM-SS 或 HH:MM, HH-MM

字符串 (String): 存放二进制信息

(5) 索引有 4 种类型: 索引、词组索引、字段特定索引、文本索引。

索引: 文本字段的每个词及其词茎被独立索引; 短语字段的每个词及其词茎以及整个字段均被索引; 整数、数值、日期和时间字段系按字段号而非内容进行索引; 字符串 (String) 字段的内容不能索引。

词组索引: 仅适用于短语字段, 只对词及词茎索引, 而不对整个字段索引。

字段特定索引: 仅用于短语字段和文本字段。短语字段或文本字段有时含有的常用词有特殊意义, 例如机构的缩写: THE、IBM 等, 除正常索引外, 还单独对此字段做索引。

文本索引: 仅用于文本字段。

4.3.9.2 修改字段

修改字段规定:

- (1) 字段号、字段名、字段类型不可修改。
- (2) 字段描述可以修改。
- (3) 子字段限制、索引类型可以修改, 修改后需重新索引。

4.3.9.3 删除字段

数据库无数据, 可以删除字段。如果数据库有数据, 不能删除字段。

4.3.10 访问权限

系统管理员和数据库管理员设定对其管理的数据库的访问权限进行授权, 被授权用户基于授权对数据库进行读访问和写访问。

对于系统管理员管理的数据库, 被授权用户包括: 所有数据库管理员、用户管理员和一般用户。

对于数据库管理员管理的数据库, 被授权用户包括: 系统管理员、其他所有数据库管理员、所有用户管理员和一般用户。

授权分为三级: 数据库级、字段级、记录级。授权规定见表 4.1。

表 4.1 授权规定

种类	功能	说明
数据库级	对数据库进行授权管理	授权用户全权读访问和写访问数据库： 1 个数据库、多个数据库或全部数据库；
字段级	对数据库的字段进行授权管理	授权用户读访问和写访问数据库的字段： 1 个字段、多个字段或全部字段；
记录级	对数据库记录的内容进行授权管理，包括：字、词、词组、数字、日期、时间、字母、字符等	授权用户读访问和写访问数据库的记录： 只能读访问和写访问含有特定内容的记录，或不含有特定内容的记录。

4.3.10.1 数据库级授权

数据库级授权是由单个数据库管理员、多个数据库管理员或所有数据库管理员授权用户，允许用户读访问和/或写访问由数据库管理员管理的 1 个数据库、多个数据库或全部数据库。

被授权用户可以是系统管理员、数据库管理员、用户管理员或一般用户。

被授权用户可以只有读访问权限；具有写访问权限的前提是必须具有读访问权限。

4.3.10.2 字段级授权

字段级授权是由单个数据库管理员、多个数据库管理员或所有数据库管理员授权用户，允许用户读访问和/或写访问由数据库管理员管理的 1 个数据库、多个数据库或全部数据库中的 1 个字段、多个字段或全部字段，见图 4.7。

被授权用户可以是系统管理员、数据库管理员、用户管理员或一般用户。

被授权用户可以只有读访问权限；具有写访问权限的前提是必须具有读访问权限。

4.3.10.3 记录级授权

记录级授权是由单个数据库管理员、多个数据库管理员或所有数据库管理员对数据库记录的内容进行授权，允许用户读访问和/或写访问由数据库管理员管理的

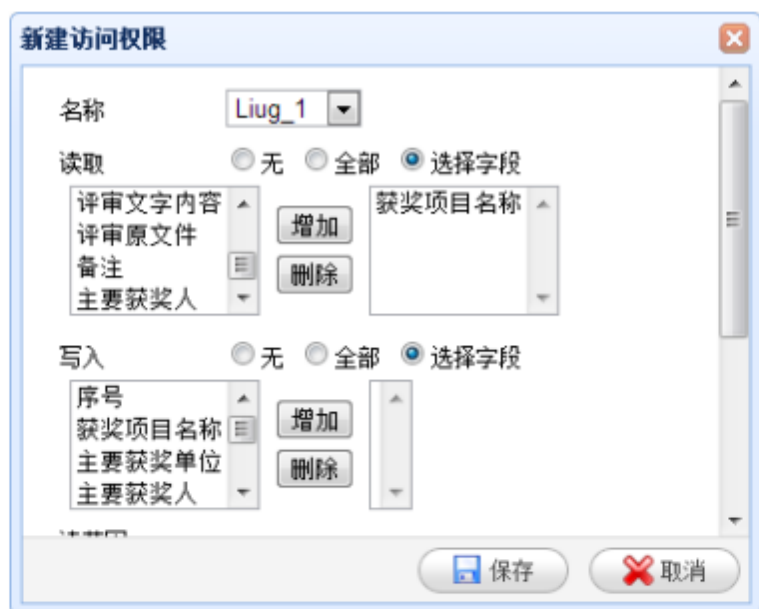


图 4.7 字段级授权对话框

1 个数据库、多个数据库或全部数据库中的全部字段或部分字段含有特定内容的记录，或不含有特定内容的记录。

被授权用户可以是系统管理员、数据库管理员、用户管理员或一般用户。

被授权用户可以只有读访问权限；具有写访问权限的前提是必须有读访问权限。

记录级授权还可分为全部字段的记录级授权(图 4.8)和部分字段的记录级授权(图 4.9)。

(1) 全部字段的记录级授权



图 4.8 记录级授权对话框

(2) 部分字段的记录级授权



图 4.9 记录级授权对话框

4.3.11 转移数据库管理权限

转移数据库管理权限只在系统管理员与数据库管理员之间，以及数据库管理员之间进行。转移数据库管理权限规定见表 4.2。

表 4.2 转移数据库管理权限规定

转移者	原所有者	新所有者
系统管理员	系统管理员	数据库管理员
数据库管理员	数据库管理员	系统管理员
数据库管理员 1	数据库管理员 1	数据库管理员 2

4.4 数据管理系统

数据管理用户界面有 6 项：档案导入、数据录入、Excel 导入、合并数据库、拆分数据库、检索编辑。

档案导入	数据录入	Excel 导入	合并数据库	拆分数据库	检索编辑
------	------	----------	-------	-------	------

4.4.1 档案导入

将数据分别写入 csv 或 Excel 表各字段列，将电子文件放入一个文件夹。分别
从 csv 或 Excel 和文件夹向数据库同时批量装入 csv 或 excel 的数据和文件夹中的
文件，见图 4.10。

国家档案局官网
WWW.SAAC.GOV.CN



图 4.10 档案导入对话框

4.4.2 数据录入

4.4.2.1 录入单个记录

该方法用于逐个录入单个记录，主要包括：

- (1) 在每个字段名后的子字段输入框中输入数据，每个框内只能输入一个子字段。
- (2) 如果一个字段有多个子字段，点击图标 ，增加子字段输入框。
- (3) 在原格式文件框，选择一个文件。

录入单个记录见图 4.11。

图 4.11 录入单个记录对话框

4.4.2.2 文件批量导入

文件批量导入仅用于导入电子文件，主要包括：

(1) 数据库只有 4 个字段：词组（存储路径和文件名）、整数（文件大小）、文本（字符内容）和二进制型（原格式文件）。例如：

TITLE_PATH:1 PHRASE COM(存储路径和文件名),

TITLE_SIZE:2 INTEGER COM(文件大小),

TITLE_TEXT:3 TEXT ORIG COM(字符内容),

TITLE:4 STRING NOX COM(原格式文件)

(2) 选择批量录入电子文件的方法有：

方式 1：最底层文件夹中的多个文件；

方式 2：最低层文件夹或一个多层文件夹；

方式 3：多个最低层文件夹或多个多层文件夹；

方式 4：多个文件和多个多层文件夹。

文件批量导入见图 4.12。

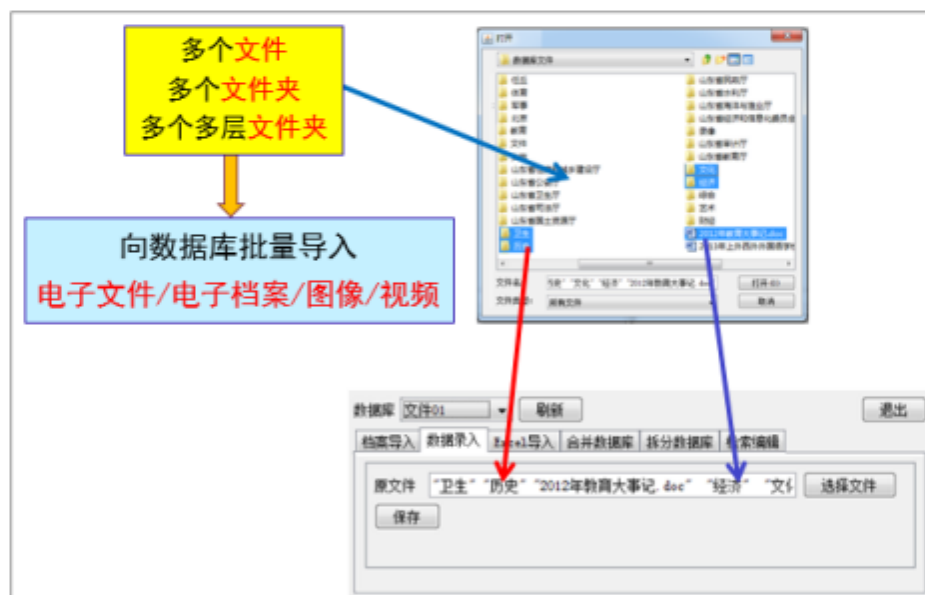


图 4.12 文件批量导入对话框

4.4.3 Excel 导入

Excel 导入是利用 csv 和 excel 表录入包含各种字段的数据和文件，见图 4.13。

将数据分别写入 csv 或 excel 表各字段列，将文件路径和文件名写入文件名字段。

A	B	C	D	E	F
项目编号	获奖项目名称	主要获奖单位	主要获奖人	推荐部门	评审材料名称
2013/1/1	国家电子文件支撑平台系统建设	国家档案局档案科学技术研究所、国家档案局技术部、国家档案局馆室司、国家档案局信息中心	李和平、马淑桂、冯丽伟、刘晓光、周萌、郝晨辉、黄静涛、徐亮	国家档案局档案科学技术研究所	d:\国家档案局\2012年度国家档案局优秀科技成果奖.doc
2013/1/2	国家电网公司档案馆数据挖掘智能化管	国家电网公司档案馆、国网中兴有限公司	曹世强、张立军、魏森、施永利、廖菲菲、马唯	国家电网公司	d:\国家档案局\国家档案局科技项目计划任务书.doc

数据库 奖13 刷新

档案导入 数据录入 Excel导入 合并数据库 拆分数据库 检索编辑

请确保Excel文件和所选数据库的结构是一致的!

Excel文件: jiang13.xls 选择文件

导入

图 4.13 Excel 导入对话框

4.4.4 合并数据库

将多个结构相同的数据库合并到一个数据库，见图 4.14。

说明：记录号重新排序。例如：A04 数据库保留原记录号，其它数据库 A0011、A031、A05 按照合并的先后次序进行记录号排序。

数据库 A04 刷新

档案导入 数据录入 Excel导入 合并数据库 拆分数据库 检索编辑

请选择将被合并到“A04”的数据库

☐ A0001
☐ 文件001
☒ A0011
☐ 视频12
☐ A01

☒ A031
☐ 刷新32
☐ A04
☒ A05
☐ A06

合并

图 4.14 数据库合并对话框

4.4.5 拆分数据库

4.4.5.1 按记录号拆分数据库

按照记录号将数据库拆分为多个子数据库，见图 4.15。

说明：

- (1) 子数据库名由英文字母、数字和下划线组成，须由英文字母开头。
- (2) 描述使用的字符不限，可使用中文。
- (3) 生成的子数据库结构和存储位置与原数据库相同。



图 4.15 按记录号拆分数据库对话框

4.4.5.2 按容量拆分数据库

按照数据库容量拆分数据库，见图 4.16。

(1) 子数据库名由原数据库、下横线_和序号组成，例如：原数据库名是 DZYH，拆分后形成的子数据库是 DZYH_1、DZYH_2、DZYH_3、……。

(2) 生成的子数据库结构和存储位置与原数据库相同。

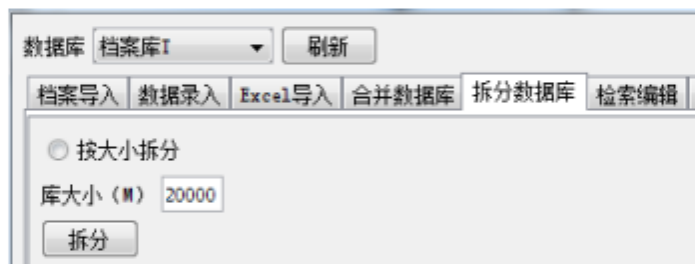


图 4.16 按容量拆分数据库对话框

4.4.5.3 按内容拆分数据库

见 4.4.8.1 输出记录到新建数据库。

4.4.6 检索编辑

4.4.6.1 增加记录内容

增加记录内容包括：

- (1) 在原有字段内容的基础上增加内容；
- (2) 增加电子文件；
- (3) 增加从电子文件中抽取出来的字符；
- (4) 增加新的字段和内容。

4.4.6.2 修改记录内容

修改记录内容包括：

- (1) 修改字段的内容;
- (2) 修改从电子文件中抽取出来的字符。

4.4.6.3 替换记录内容

替换记录内容包括:

- (1) 替换字段的内容;
- (2) 替换从电子文件中抽取出来的字符;
- (3) 替换数据库中的电子文件。

4.4.6.4 删除记录内容

删除记录内容包括:

- (1) 删除字段的内容;
- (2) 删除从电子文件中抽取出来的字符;
- (3) 删除数据库中的电子文件。

4.4.6.5 删除记录

从数据库中删除记录包括: 删除全部记录和删除部分记录。

删除部分记录分为: 按记录号顺序删除和选择性删除。

4.4.7 输出记录

4.4.7.1 输出记录到新建数据库

按检索结果输出记录到新建数据库, 包括: 各字段的内容、文件名、文件大小, 从文件中抽取出来的字符和电子文件, 但不保留记录在原数据库中的记录号, 而是重新排序, 见图 4.17。新建数据库的存储路径与原数据库的存储路径相同。

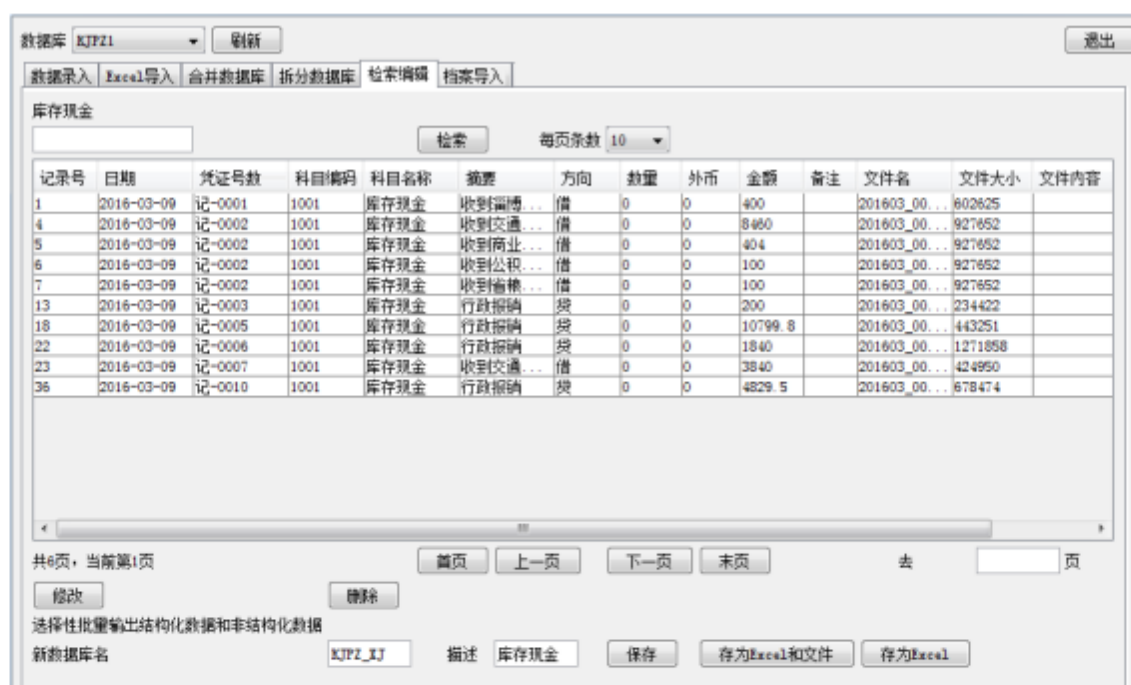


图 4.17 输出记录到新建数据库

4.4.7.2 输出记录到 Excel 和文件夹

检索后从数据库输出记录，著录信息的字段内容和电子文件原文名存入 csv 或 Excel 表，电子文件存入文件夹，见图 4.18。

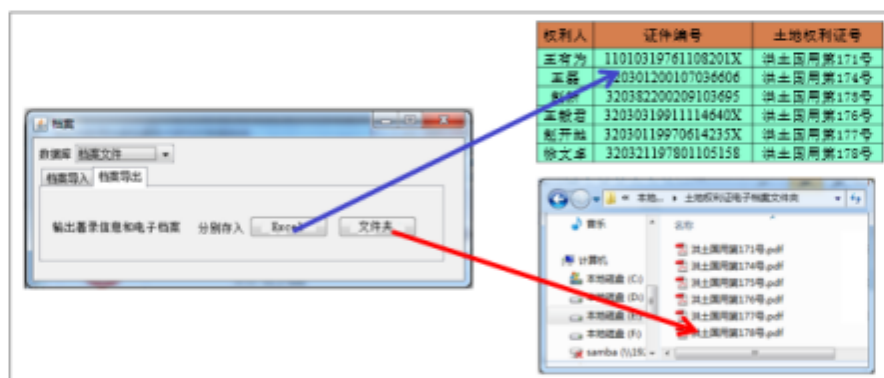


图 4.18 输出记录到 Excel 和文件夹

4.4.7.3 输出记录到 Excel

检索后从数据库输出记录的著录信息的字段内容和电子文件原文名存入 Excel。

4.5 检索系统

4.5.1 检索授权

系统管理员和数据库管理员对其管理的数据库的检索进行授权，被授权用户基

于被授权检索的条件进行检索。检索包括：数据库无限制检索、数据库限制检索、字段无限制检索、字段限制检索、内容无限制检索、内容限制检索。授权检索规定见表 4.3。

表 4.3 授权检索规定

种类	功能	说明
数据库无限制检索	不限制对数据库的检索	允许检索全部数据库； 显示全部数据库；
数据库限制检索	对数据库检索进行限制	不允许检索全部数据库，允许检索部分数据库； 不显示全部数据库，仅显示部分数据库；
字段无限制检索	不限制对字段（栏目）的检索	允许检索全部字段； 显示全部字段；
字段限制检索	对字段（栏目）检索进行限制	不允许检索全部字段，允许检索部分字段； 不显示全部字段，仅显示部分字段；
内容无限制检索	不限制对记录内容的检索	允许检索全部记录； 显示全部记录；
内容限制检索	对记录内容检索进行限制	不允许检索全部记录，允许检索部分记录； 不显示全部记录，仅显示部分记录。

4.5.1.1 数据库无限制检索

数据库无限制检索是由系统管理员、单个数据库管理员、多个数据库管理员或所有数据库管理员授权用户，允许用户检索和读取由系统管理员和数据库管理员管理的全部数据库。被授权用户可以是系统管理员、数据库管理员、用户管理员或一般用户。

例：数据库管理员 SONG 管理 3 个数据库，数据库管理员 ZHANG 管理 2 个数据库。数据库管理员 SONG 和 ZHANG 授权用户 QIAN，允许用户 QIAN 检索和读取由数据库管理员 SONG 和 ZHANG 管理的全部 5 个数据库。

4.5.1.2 数据库限制检索

数据库限制检索是由系统管理员、单个数据库管理员、多个数据库管理员或所有数据库管理员授权用户，允许用户检索和读取由数据库管理员管理的部分数据库。被授权用户可以是系统管理员、数据库管理员、用户管理员或一般用户。

例：数据库管理员 SONG 管理 3 个数据库，数据库管理员 ZHANG 管理 2 个数据库。数据库管理员 SONG 和 ZHANG 授权用户 QIAN，允许用户 QIAN 检索和读取由数据库管理员 SONG 和 ZHANG 管理的 5 个数据库中的 3 个数据库，其它 2 个数据库不可检索，在数据库列表中也不显示。

4.5.1.3 字段无限制检索

字段无限制检索是由系统管理员、单个数据库管理员、多个数据库管理员或所有数据库管理员授权用户，允许用户检索和读取由系统管理员和数据库管理员管理的单个数据库、部分数据库或全部数据库的全部字段。被授权用户可以是系统管理员、数据库管理员、用户管理员或一般用户。

例：数据库管理员 SONG 管理的数据库“奖 38”有 10 个字段，允许用户 QIAN 检索和读取数据库“奖 38”的全部 10 个字段。

4.5.1.4 字段限制检索

字段限制检索是由系统管理员、单个数据库管理员、多个数据库管理员或所有数据库管理员授权用户，允许用户检索和读取由系统管理员和数据库管理员管理的单个数据库、部分数据库或全部数据库的部分字段。被授权用户可以是系统管理员、数据库管理员、用户管理员或一般用户。

例：数据库管理员 SONG 管理的数据库“奖 38”有 10 个字段，授权用户 QIAN 检索和读取数据库“奖 38”的 4 个字段：获奖项目名称、主要获奖人、评审材料名称、评审文字内容（文本字段）。被限制的字段包括：评审原文件（字符串字段）。

4.5.1.5 内容无限制检索

内容无限制检索分为：字段无限制检索条件下的内容无限制检索和字段限制检索条件下的内容无限制检索。

(1) 字段无限制检索条件下的内容无限制检索

字段无限制检索条件下的内容无限制检索是由系统管理员、单个数据库管理员、多个数据库管理员或所有数据库管理员授权用户，允许用户检索和读取由系统管理员和数据库管理员管理的单个数据库、部分数据库或全部数据库的所有字

段的全部内容。被授权用户可以是系统管理员、数据库管理员、用户管理员或一般用户。

例：数据库管理员 SONG 管理的数据库“奖 38”有 10 个字段，允许用户 QIAN 检索和读取数据库“奖 38”的全部 10 个字段的所有内容。

(2) 字段限制检索条件下的内容无限制检索

字段限制检索条件下的内容无限制检索是由系统管理员、单个数据库管理员、多个数据库管理员或所有数据库管理员授权用户，允许用户检索和读取由系统管理员和数据库管理员管理的单个数据库、部分数据库或全部数据库的部分字段的全部内容。被授权用户可以是系统管理员、数据库管理员、用户管理员或一般用户。

例：数据库管理员 SONG 管理的数据库“奖 38”有 10 个字段，允许用户 QIAN 检索和读取数据库“奖 38”的 3 个字段的所有内容。

4.5.1.6 内容限制检索

内容限制检索分为：字段无限制检索条件下的内容限制检索和字段限制检索条件下的内容限制检索。

(1) 字段无限制检索条件下的内容限制检索

字段无限制检索条件下的内容限制检索是由系统管理员、单个数据库管理员、多个数据库管理员或所有数据库管理员授权用户，允许用户检索和读取由系统管理员和数据库管理员管理的单个数据库、部分数据库或全部数据库的所有字段的部分内容。被授权用户可以是系统管理员、数据库管理员、用户管理员或一般用户。

a) 只允许用户检索和读取包含特定内容的记录

例：数据库管理员 SONG 管理的数据库“奖 38”有 10 个字段，允许用户 QIAN 检索和读取数据库“奖 38”的全部 10 个字段中包含特定内容“国家电网”的记录。

b) 只允许用户检索和读取不包含特定内容的记录

例：数据库管理员 SONG 管理的数据库“奖 38”有 10 个字段，允许用户 QIAN 检索和读取数据库“奖 38”的全部 10 个字段中不包含特定内容“国家电网”的记录。

(2) 字段限制检索条件下的内容限制检索

字段限制检索条件下的内容限制检索是由系统管理员、单个数据库管理员、多

个数据库管理员或所有数据库管理员授权用户，允许用户检索和读取由系统管理员和数据库管理员管理的单个数据库、部分数据库或全部数据库的部分字段的部分内容。被授权用户可以是系统管理员、数据库管理员、用户管理员或一般用户。

a) 只允许用户检索和读取包含特定内容的记录

例：数据库管理员 SONG 管理的数据库“奖 38”有 10 个字段，允许用户 QIAN 检索和读取数据库“奖 38”的 3 个字段中包含特定内容“国家电网”的记录。

b) 只允许用户检索和读取不包含特定内容的记录

例：数据库管理员 SONG 管理的数据库“奖 38”有 10 个字段，允许用户 QIAN 检索和读取数据库“奖 38”的 3 个字段中不包含特定内容“国家电网”的记录。

4.5.2 数据库选择检索

数据库选择检索是用户在被系统管理员、单个数据库管理员、多个数据库管理员或所有数据库管理员授权的一个数据库或一个以上的数据库中选择数据库进行检索。被授权用户可以是系统管理员、数据库管理员、用户管理员或一般用户。

4.5.3 检索条件选择

检索条件选择有三种方式：字段类型、记录号和字段名。

采用字段类型检索适用于单个数据库、同构多库和异构多库。由于异构多库中每个数据库的字段名不尽相同或完全不同，采用字段名，就需要列出所有数据库的全部字段名，所以，异构多库一般采用字段类型，而不采用字段名。可用于字段类型检索的字段类型有：phrase（词组）和 text（文本）。date（日期）、time（时间）、整数（Integer）、数值（Number）必须给出数值、日期、时间的实际字段名称，而非字段类型。

采用记录号适用于单个数据库、同构多库和异构多库。

采用字段名适用于单个数据库和同构多库。

4.5.4 二次检索

二次检索是在上一次检索结果基础上进一步查找。

4.5.5 间接检索

所谓间接检索，是用数据库检索命中记录的一个指定字段的词汇作为检索词，在另一个数据库进行检索。也就是说，先对一个源数据库进行检索，获取一组命中记录，然后用该组记录中的一个指定字段的全部词作为检索词，对另一目标数

数据库进行检索。这种检索方法提供了两个数据库之间的关联查找。

间接检索的过程分如下三步：

- (1) 从源数据库中查找出一组记录，以便从中抽取出检索词对目标数据库进行检索；
- (2) 从命中记录中的一个词组 **Phrase** 字段（作为源字段）提取出内容；
- (3) 以源字段所提取的内容作为检索词在一个或多个目标数据库中的一个或多个字段中进行查找。

主要操作步骤

- (1) 打开目标数据库：**base target_databse**。
- (2) 定义虚拟字段：用如下命令定义一个指向源数据库、源字段的虚拟字段：

```
de map vfn=source_database.fieldname
```

vfn 对任何数据库都不存在的虚拟字段

map 定位符

- (3) 将这个虚拟字段指向一个源数据库 **source_database**，隐含的是在源数据库中的 **Phrase (ph)** 字段、**Text (te)** 字段查找，同时指定从源数据库中的一个字段 **fieldname** 中取出词汇，用作检索词，在目标数据库 **target_database** 中的 **te**、**ph** 字段中检索。上述定义会自动转变为：

```
Define map vfn=TE+PH:Source_base.fieldname:TE+PH
```

第一个和第二个 **TE+PH** 可以只有 **PH** 或 **TE**，隐含的是 **TE+PH**；

fieldname 必须是源数据库中的一个 **Phrase** 字段名，不能用 **PH** 字段类型。

- (4) 用 **find** 命令通过对虚拟字段的检索进行目标数据库的间接查找：**F vfn (search_term)**。

间接检索表面看是对虚拟字段的检索，实际上是按照给定的检索项 **search_term** 在源数据库的所有 **Phrase** 和 **Text** 字段中进行匹配查找，取得命中记录，然后从各个命中记录的指定字段 **fieldname** 中取出词汇，用这些词汇作为检索词，在目标数据库 **target_database** 中的 **Phrase** 和 **Text** 字段中进行查找。检索项 **search_term** 可以是一个或多个检索条件的任意组合。

4.5.6 多数据库检索

多数据库检索分为：同构多库检索、异构多库检索。

(1) 同构多库检索：同时检索结构相同的多个数据库。

(2) 异构多库检索：同时检索结构不同的多个数据库。

4.5.7 异构类型数据库检索

4.5.7.1 基于磁盘的异构类型数据库检索

关系数据库和非关系型数据库都建立在磁盘上，将结构化数据导入关系数据库，将非结构化数据导入非关系数据库，通过跳转程序连接关系数据库与非关系数据库，形成关系数据库—非关系数据库系统。根据关系数据库中的结构化数据查询结果，可输出相连接的非关系数据库中的非结构化数据，见图 4.19。

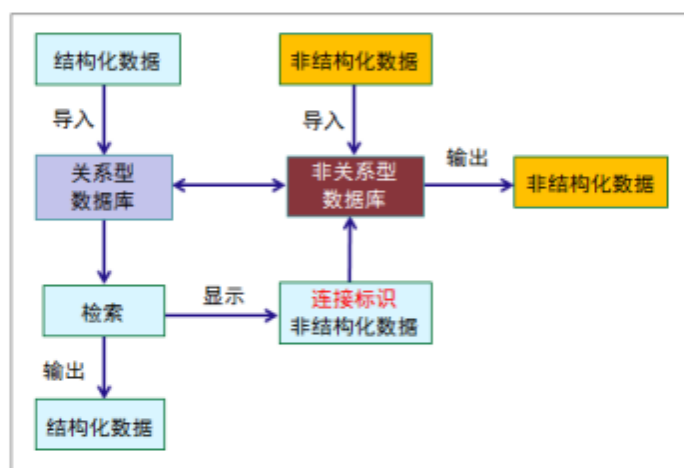


图 4.19 基于磁盘的异构类型数据库检索示意图

4.5.7.2 基于光盘的异构类型数据库检索

关系数据库建立在磁盘上，非关系型数据库建立在光盘库中的光盘上，将结构化数据导入关系数据库，将非结构化数据导入非关系数据库，通过跳转程序连接关系数据库与非关系数据库，形成关系数据库——非关系型数据库——光盘库系统。根据关系数据库中的结构化数据查询结果，输出相连接的光盘库中光盘上的非关系数据库中的非结构化数据，见图 4.20。

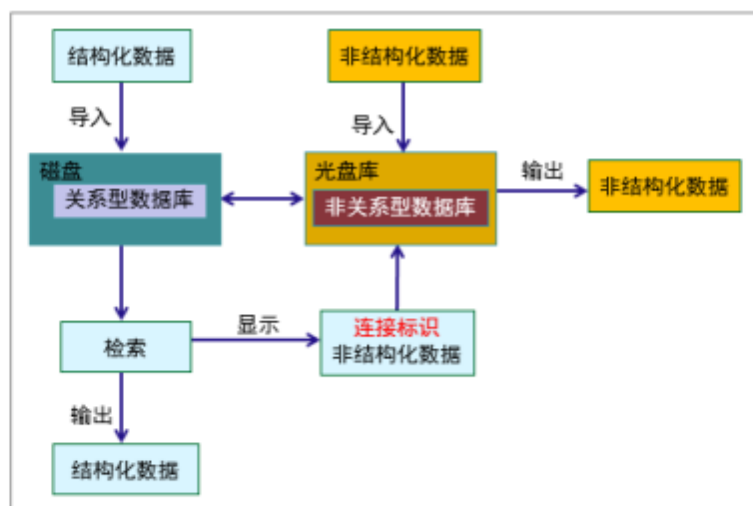


图 4.20 基于光盘的异构类型数据库检索示意图

4.5.8 光盘内容全文检索

将数据库建立在光盘上，将这种数据库光盘放入与服务器直接相连的光盘驱动器或光盘库中，可对一个数据库光盘中的数据库或多个数据库光盘中的数据库进行连续的全文检索，包括二次检索、布尔检索等高级检索，不再需要将数据从数据库光盘中的数据库先还原到磁盘存储装置中。利用数据库拆分和多盘检索，能够将大数据库拆分为多个子数据库，将子数据库分别存储在多张光盘上进行检索。

根据检索的光盘数量，分为单盘检索和多盘检索。根据检索的光盘数量和光盘上数据库的数量，又可分为单盘多数据库检索和多盘多数据库检索。

4.5.8.1 单盘单库检索

单盘单库检索是指对单张光盘上的单个数据库进行检索。

4.5.8.2 多盘多库检索

多盘多库检索是指对多张光盘上的多个数据库进行检索，需要使用光盘库存储光盘。一张光盘上可以有一个数据库，也可以有多个数据库。多库检索可以是同构多库检索，也可以是异构多库检索。

5 研究意义

5.1 学术价值

5.1.1 非关系数据库存储电子档案的新思路和新方法的创立,推进了电子档案存储、检索和利用技术的进步

运用非关系数据库技术和数据库双核存储技术,将各种格式的电子档案全部装入数据库,不采用存储路径的方式将电子档案挂接在数据库上的方法,也不采用将需要挂接的电子档案嵌入数据库的方法,保证了电子档案的完整性和安全性,彻底改变了电子档案挂接或嵌入数据库的传统模式,大幅度提高了电子档案存储的安全性和完整性,推进了电子档案存储、检索和利用的技术进步。

5.1.2 异构类型数据库存储电子档案新方法的创立,发挥了关系数据库和非关系数据库各自的优势,提高了电子档案数据存储的效率

运用异构类型数据库存储技术,充分利用关系数据库和非关系数据库的优势和特点,适应用户的使用习惯,将著录信息存储在关系数据库,将电子档案原文存储在非关系数据库,不再将电子档案原文挂接在关系数据库上,使得所有的数据都存储在数据库中,解决了电子档案中的不同结构类型数据无法结合数据库类型进行数据存储的技术问题,保证了电子档案的完整性;对于已有的关系数据库,将挂接在关系数据库上的电子档案原文转储到非关系数据库,解决了电子档案中的结构化数据与非结构化数据通过单一类型数据库进行数据处理时导致数据处理性能下降的技术问题,大幅度提高了关系数据库的存储能力和安全性,为关系数据库的有效利用开辟了新的途径和方法。

5.1.3 数据库拆分和合并技术的研发,提供了电子档案分类和汇聚整合的新方法

存储电子档案的非关系数据库的记录包括著录信息(结构化数据和非结构化数据)和电子档案原文(非结构化数据)。数据库拆分是以记录为单元拆分,同时将著录信息和电子档案原文拆分出来。可按记录数、容量大小、档案的内在联系等拆分数据库。利用非关系数据库拆分和合并对电子档案进行分类和汇聚整合,使得电子档案的存储和管理具有足够的灵活性和广泛的适用性,促进电子档案存储模式的转变和技术进步,大幅度提高了电子档案的应用广度和深度。

5.1.4 基于光盘的非关系数据库存储电子档案新技术的研发，是光盘存储电子档案并进行全文检索方式的革新

不同于现有的将数据库数据备份在光盘上的技术，利用自主研发的数据库拆分、合并技术，将非关系数据库建立和存储在光盘上，实现了对光盘上建立和存储的数据库中的电子档案内容进行全文检索；利用多（光）盘多库检索技术，实现了光盘库中数据库的全文检索，且速度达到毫秒级，调取电子档案达到秒级，完全达到了实用的水平。

5.2 实际应用价值

5.2.1 《基于非关系数据库的电子档案存储规范》的研制，为电子档案存储提供了新方法

《基于非关系数据库的电子档案存储规范》规定了采用非关系数据库在磁盘和蓝光光盘上存储电子档案和档案数字副本的方法，即基于非关系数据库技术、数据库双核存储技术、非关系数据库拆分和合并技术，将著录信息和电子档案原文存储在非关系数据库中。该规范改变了电子档案原文挂接或嵌入数据库的传统模式，适用于非结构化、半结构化和结构化档案数据的存储。

5.2.2 基于非关系数据库的电子档案存储系统为实现“建得起，用得起，可持续，可落地”的电子档案的存储、检索和利用系统提供了坚实基础和有效工具，具有良好的推广应用前景

课题组研制的基于非关系数据库的电子档案存储系统架构设计科学，用户管理、数据库管理和检索采用 B/S 方式，数据管理采用 C/S 方式，配置灵活、适用性强，具有强大的电子档案存储、管理、汇聚整合和检索功能。自主研发的应用软件用户界面友好，易于掌握和操作。

与关系数据库表结构的设计不同，非关系数据库可以创建不同类型的非结构化数据的字段，从而突破了关系数据库严格的表结构，解决了由于关系数据库模型简单，不易表达复杂嵌套数据结构的问题，能够存储各种类型的电子档案。该系统的显著特点之一是大大降低了数据库设计和建立的技术难度，非数据库专业领域的用户能够按照需求合理、快捷地设计和建立数据库，还可根据实际情况，修改数据库的结构和字段，大幅度缩短了数据库设计和建立的时间。

该系统的数据库管理模式不同于关系数据库，有利于非数据库专业领域的用户

管理数据库。单台服务器可管理数百个数据库，不同的数据库可分属不同的数据库管理员管理，数据库管理员的数量无限制。数据库访问授权分为三级：数据库级、字段级、记录（内容）级，由分管的数据库管理进行访问授权。被授权用户限制访问分为三级：数据库级、字段级、记录（内容）级，被限制检索的部分不显示，即：被限制检索的数据库、被限制检索的字段、被限制检索的记录（内容）不显示。

该系统的显著特点之一是具有高效实用的数据导入功能和数据输出功能，能够将各种格式的电子文件批量混合导入数据库，将著录信息和各种格式的电子档案按照一定的匹配规律批量混合导入数据库，且不需对电子档案进行预整理，也不需标注电子档案存储路径，大大降低了数据准备的工作量，大幅度提高了数据导入的效率；能够批量将数据库中记录的著录信息和电子文件原文名输出到 csv 或 Excel 表，电子档案原文输出到文件夹，大幅度提高了数据输出的效率。

该系统的突出特点是电子档案的汇聚整合功能，是将各种格式的海量电子档案批量混合导入数据库，通过检索或知识挖掘对数据库中的电子档案进行分类，并将分类获取的电子档案从原数据库拆分输出到子数据库，然后对多个同类的子数据库进行合并，形成类别数据库。电子档案的汇聚整合功能可以广泛用于电子档案的归类整理，大幅度降低了电子档案分类存储的复杂程度和劳动强度，提高了电子档案管理的科学性和效率。

该系统能够对磁盘存储装置和光盘存储装置上数据库中的电子档案进行统一管理，充分利用磁盘存储装置和光盘存储装置各自的优势对电子档案进行分级存储，大幅度降低存储和管理成本，并能够采用同构多库检索、异构多库检索、多（光）盘多库检索技术，对磁盘存储装置和光盘存储装置上的数据库进行全文检索。

课题组采用模拟数据对系统和应用软件进行压力测试。测试结果表明，单个数据库能够存储 10 亿（条）级著录信息和千万（个）级电子档案，使用单台 32GB 内存服务器能够管理数百个数据库的千亿（条）级著录信息和 10 亿（个）级电子档案，容量达到 PB 级，全文检索速度达到毫秒级；著录信息导入数据库的速度达到千（条）级/秒，电子档案的导入速度达到 20MB/秒级。

该系统能够充分利用关系数据库和非关系数据库的优势和特点，适应用户的使

用习惯，在基于非关系数据库的电子档案存储系统的基础上，建立异构类型数据库存储系统，有效利用现有的关系数据库存储电子档案的资源，大幅度提高了关系数据库的存储能力和安全性。

本课题研制的基于非关系数据库的电子档案存储系统稳定，大幅度降低了对软硬件资源的配置要求和维护成本，非计算机和数据库专业领域的用户能够在短时间内熟练掌握该系统，适合各级档案部门采用非关系数据库技术存储、管理和利用电子档案，为实现“建得起，用得起，可持续，可落地”的电子档案的存储、检索和利用系统提供了坚实基础和有效工具，具有良好的推广应用前景。

5.3 经济效益

本课题的研究成果将大幅度减少软硬件成本、管理成本、维护成本、时间成本，降低能源消耗；大幅度减少电子档案数据库迁移和备份的工作量和费用，提高工作效率。

参考文献

- [1] 刘家真, 文件保存格式与 PDF 文档[J], 档案学研究, 2002(2):46-51
- [2] 李俊仁, 朱晓钟. 电子档案存储方式的选择[J]. 硅谷, 2010(9):111-111
- [3] 史玉秋. 电子档案的长期存储选择探析[J]. 赤子, 2013[21]:234-234
- [4] 杨晶. 数字档案馆信息检索系统的研究与实现[D]. 北京交通大学, 2007
- [5] 苏禅. 非关系型数据库及在文档库管理平台中的研究与应用[D]. 南昌大学, 2014
- [6] 田伟, 韩海涛. 一种基于非关系型数据库的大数据档案资源处理策略[J]. 兰台世界, 2015, No.484(26):21-22
- [7] 程传鹏. 基于 Trip 数据库的档案管理系统的设计与实现[J]. 中原工学院学报, 2012, 23(1):44-48
- [8] 江西省国家综合档案馆馆藏专业档案目录数据库结构规范

附件 1

基于非关系数据库的电子档案存储规范

1 范围

本标准规定了采用非关系数据库在磁盘和蓝光光盘上存储电子档案和档案数字副本的方法。

本标准适用于非结构化、半结构化和结构化档案数据的存储。

2 术语和定义

下列术语和定义适用于本文件。

2.1 数据库 database

按照概念结构组织的数据的汇集，它描述这些数据的特征及与数据对应的实体间的关系并支持一个或多个应用领域。

[GB/T 5271.1-2000, 定义 01.08.05]

2.2 关系数据库 relational database

数据按关系模型来组织的数据库。

[GB/T 5271.1-2000, 定义 17.04.05]

2.3 非关系数据库 non-relational database

采用子字段、多值字段和变长字段机制，创建不同类型的非结构化的或任意格式的字段，以多维处理方式突破关系数据库二维表结构，实现从数据属性管理转化为内容管理的数据库。

2.4 数据库管理系统 database management system

基于硬件与软件，用于定义、建立、操纵、控制、管理和使用数据库的系统。

[GB/T 5271.1-2000, 定义 17.01.03]

2.5 关系数据库管理系统 relational database management system

用于关系数据库的数据库管理系统。

[GB/T 5271.1-2000, 定义 17.04.06]

2.6 非关系数据库管理系统 non-relational database management system

用于非关系数据库的数据库管理系统。

2.7 数据库文件组织 database file organization

对存储器中数据的安排和存取方法的实现：使之都符合特别文件及其记录的数据结构，并将此文件规定为数据库的组成部分。

[GB/T 5271.1-2000, 定义 17.03.10]

注：本存储规则提及的每一个应用数据库均包含两个独立的文件：一是用于存放原始数据信息的 BAF 文件，即主文件；二是用于存放 BAF 文件中包含的所有词条的位置信息的 BIF 文件，即索引文件。

2.8 结构化数据 structured data

能够用简短的数字、文字或统一的结构加以表示的数据。

2.9 半结构化数据 semi-structured data

具有结构性，但结构变化大，且难以用结构化数据的处理方法将其放进平面行列表格里的数据。

2.10 非结构化数据 unstructured data

不能用数字或者统一的结构表示或没有固定结构，且不能用平面行列表格结构存放的数据。

3 电子档案存储方式

3.1 概述

数据库包括关系数据库、非关系数据库，本标准采用的数据库为非关系数据库；数据库管理系统包括关系数据库管理系统、非关系数据库管理系统，本标准采用的数据库管理系统为非关系数据库管理系统。

数据库存储系统包括服务器、与服务器数据连接的磁盘存储装置和光盘存储装置。数据库管理系统、数据处理器、数据连接器安装在服务器上。数据库中设置数据库基本核和数据库扩展核（见图 1）。

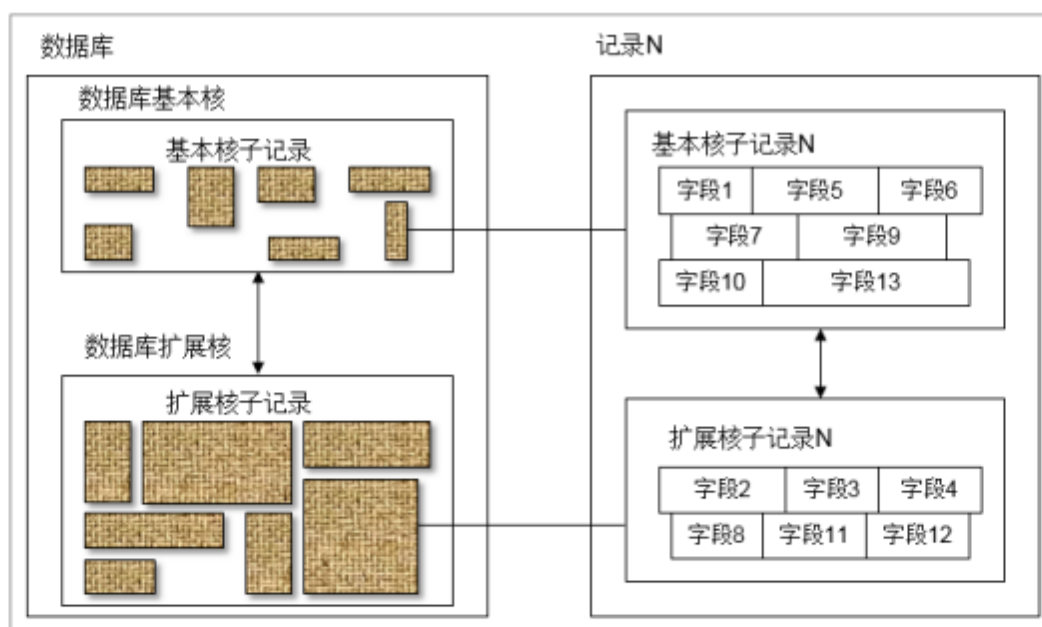


图 1 数据库存储系统的数据库文件的物理结构

数据库管理系统用于磁盘存储装置和光盘存储装置中数据库管理和数据管理；数据处理器用于配置数据库基本核的字段和数据库扩展核的字段，并将相应字段的数据分别写入数据库基本核和数据库扩展核；数据连接器用于建立数据库基本核和数据库扩展核的数据连接。数据库管理系统包括字段结构生成模块、记录索引生成模块、存储空间分配模块、存储空间状态模块、记录链接模块、记录拆分模块、库结构生成模块、双核生成模块、记录合并模块、检索词存储模块、检索词索引生成模块；数据处理器包括字段配置模块、数据录入模块；数据连接器包括双核连接模块（见图 2、图 3 和图 4）。

数据库的主文件和索引文件是两个独立的数据库文件，数据库主文件直接存储电子档案，而不是采用存储路径的方式挂接在数据库上，每一条电子档案记录由著录信息、电子档案原文名、从电子档案原文中抽取出的字符、电子档案原文组成（见图 5 和图 6）；索引文件存储电子档案记录全部词条的索引信息。

数据库主文件和索引文件可以分别存储在不同的存储介质和不同的位置。

3.2 采用磁盘的存储方式

数据库主文件和数据库索引文件都存储在磁盘上（见图 2）。

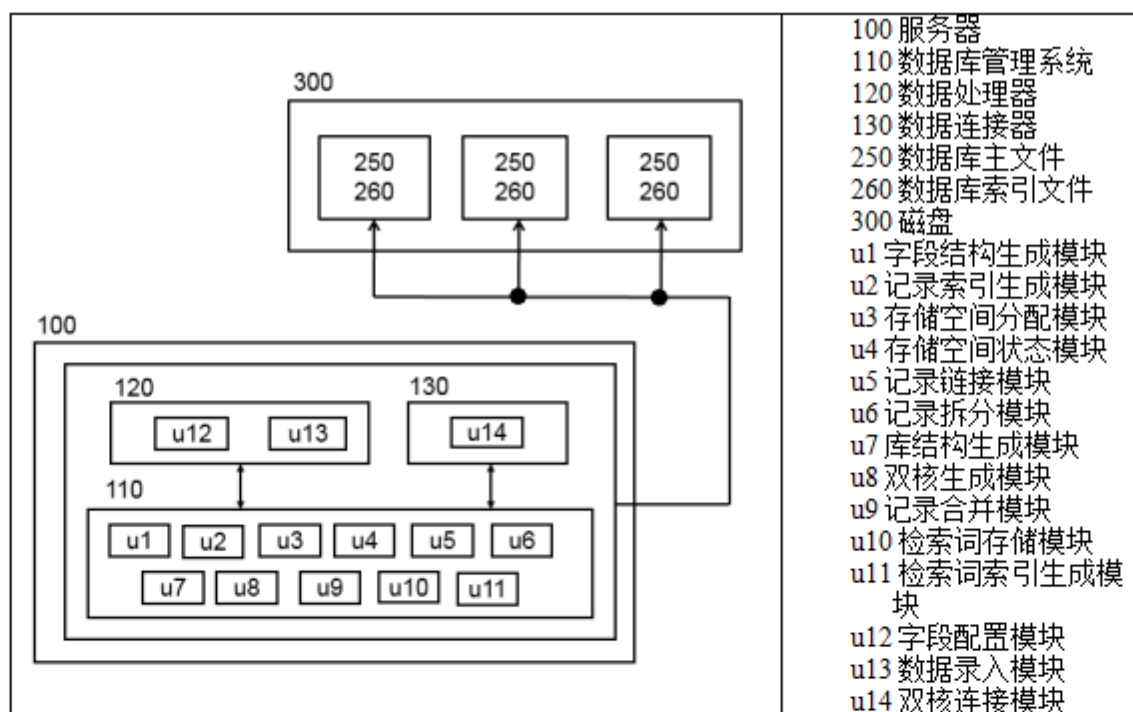


图 2 基于磁盘的数据库存储系统结构示意图

3.3 采用光盘的存储方式

数据库主文件存储在光盘上，数据库索引文件存储在磁盘（见图 3）或光盘上（见图 4）

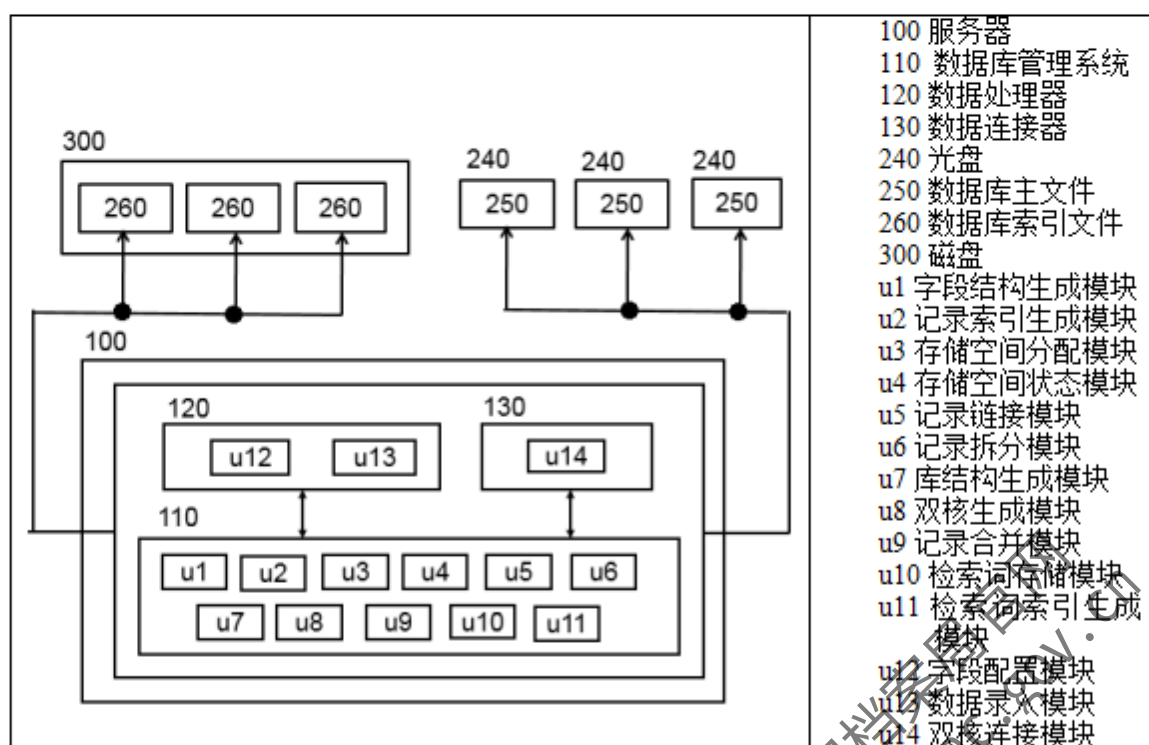


图 3 基于光盘的数据库存储系统结构示意图（索引文件在磁盘上）

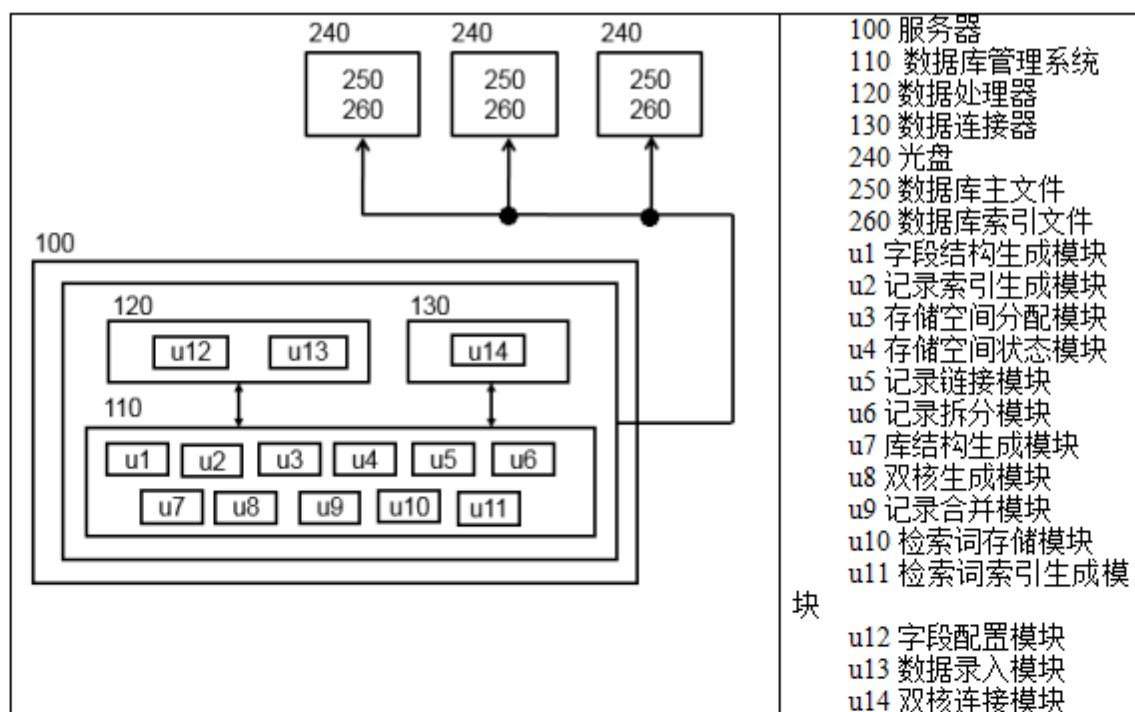


图 4 基于光盘的数据库存储系统结构示意图（索引文件在光盘上）

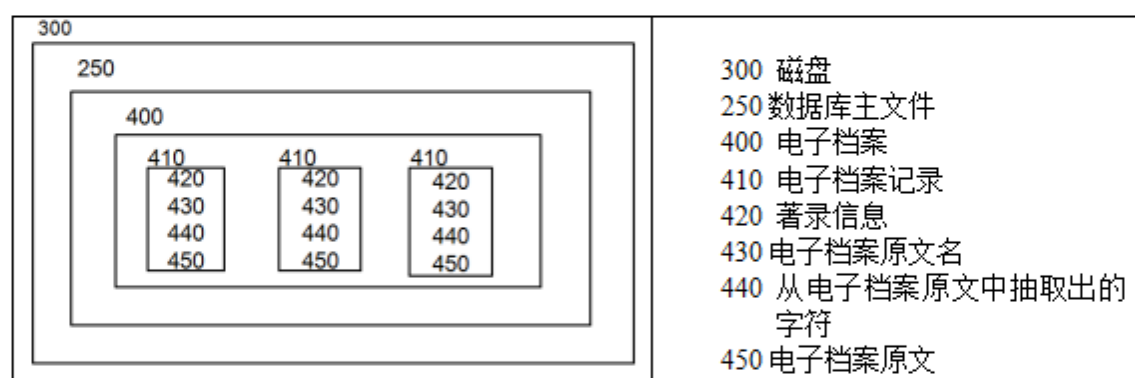


图 5 基于磁盘的数据库主文件存储电子档案结构示意图

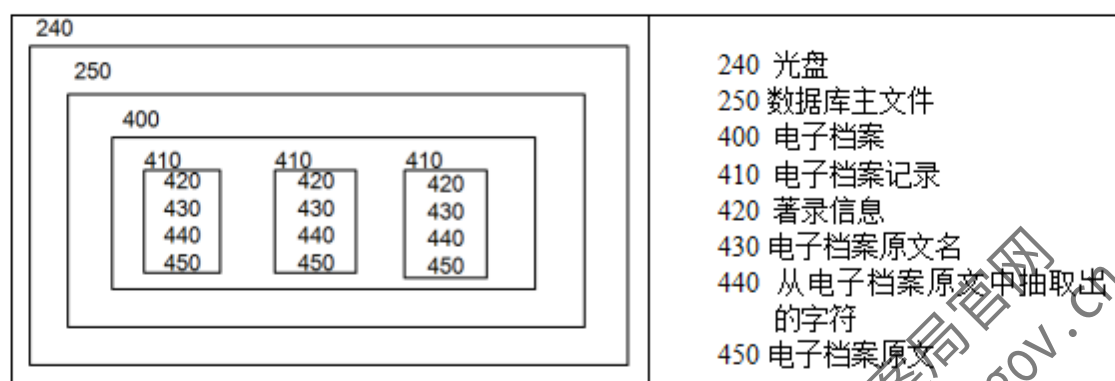


图 6 基于光盘的数据库主文件存储电子档案结构示意图

3.4 数据库管理系统

3.4.1 字段结构生成模块

读取数据库中每一条记录的字段结构信息，包括记录中各字段的数据类型、长度，并写入数据库管理系统中。

3.4.2 记录索引生成模块

记录数据库中每一条记录的索引信息，包括记录中各字段的修改时间、修改内容，并写入数据库管理系统。

3.4.3 存储空间分配模块

记录为每一条记录所分配的标准存储单元在数据库文件中的位置信息，并写入数据库管理系统。

3.4.4 存储空间状态模块

记录数据库文件中已分配标准存储单元中未使用的空间信息，并写入数据库管理系统。

3.4.5 记录链接模块

将数据库中各记录的字段结构信息、索引信息、标准存储单元的位置信息和空间信息合并，形成数据库特征数据，并写入数据库管理系统。

3.4.6 记录拆分模块

根据数据库管理系统指令进行数据库中记录的拆分，数据拆分以记录为单位进行，读取数据库特征数据，确定记录中每个字段的数据位置和数据量，标记出符合指令参数的记录，并将标记信息写入数据库管理系统。

3.4.7 库结构生成模块

将数据库的库结构形成独立数据文件，数据库管理系统根据独立数据文件在磁盘上建立相同库结构的数据库，或光盘上建立相同库结构的数据库。

3.4.8 双核生成模块

通过数据库管理系统在磁盘上的数据库中设置数据库基本核或数据库基本核的数据库文件，设置数据库扩展核或数据库扩展核的数据库文件；或在光盘上的数据库中设置数据库基本核或数据库基本核的数据库文件，设置数据库扩展核或数据库扩展核的数据库文件。

3.4.9 记录合并模块

根据数据库管理系统指令，将光盘上子数据库中的记录合并到磁盘存储装置

上的磁盘数据库中，并通过数据库管理系统生成对应记录的字段结构信息、索引信息，以及记录标准存储单元位置信息和空间信息，形成磁盘数据库的数据库特征数据。

3.4.10 检索词存储模块

用于存储包含语义信息的检索词字库，检索词至少包括字、词和数字。

3.4.11 检索词索引生成模块

根据数据库中每一条记录的索引信息，建立与标记信息对应的记录的检索词索引数据，包括检索词出现的频率和在每一条记录中的位置，并写入数据库管理系统。

3.4.12 字段配置模块

根据记录中各字段的数据类型和长度，配置数据库基本核的字段和数据库扩展核的字段，形成基本核子记录和扩展核子记录，并将字段配置信息写入数据库文件或数据库管理系统中。

3.4.13 数据录入模块

根据字段配置模块配置的数据库基本核的字段和数据库扩展核的字段，按照数据库管理系统指令，将相应字段的数据分别写入磁盘上的数据库基本核和数据库扩展核，或光盘上的数据库基本核和数据库扩展核。

3.4.14 双核连接模块

按照数据库管理系统的指令，连接数据库的基本核子记录和相应的扩展核子记录，形成完整的记录。

3.5 数据库拆分

数据库拆分是以记录为单元进行拆分。

一个数据库被拆分成若干个结构定义一致、数据完整的子数据库，同时生成相应的子数据库主文件（见图 7）。

对每个子数据库主文件单独进行索引，生成与子数据库主文件相对应的子数据库索引文件（见图 7）。

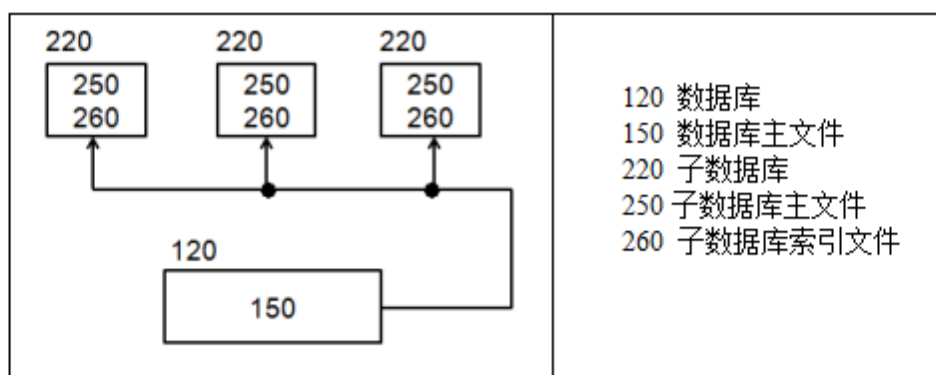


图 7 数据库拆分示意图

数据库中的各记录关联数据不需要跨子数据库存储或关联，使得数据库在存储过程中可以保持数据的完整性，每一个子数据库作为正常的数据源接受数据库管理系统的管理。

3.6 数据库合并

数据库合并是以记录为单元进行合并。

若干个结构定义一致的数据库合并生成一个大数据库，同时生成相应的大数据库文件（见图 8）。

对合并生成的大数据库主文件进行索引，生成与大数据库主文件相对应的大数据库索引文件（见图 8）。

生成的大数据库作为正常的数据源接受数据库管理系统的管理。

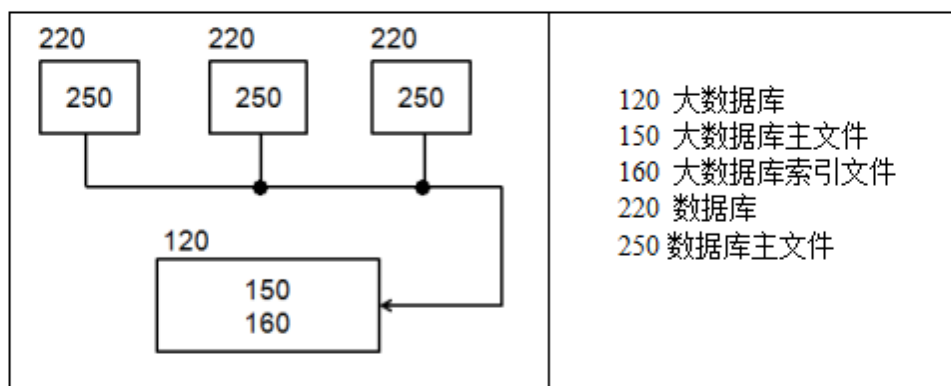


图 8 数据库合并示意图

3.7 电子档案录入导入与导出

3.7.1 电子档案录入导入

单个录入或批量导入著录信息、电子档案原文名、从电子档案原文中抽取出的字符和电子档案原文。每一条电子档案记录可容纳著录信息、电子档案原文名、从电子档案原文中抽取出的字符和电子档案原文（见图 9）。

著录信息和电子档案原文名是结构化数据，从电子档案原文中抽取出的字符

和电子档案原文是非结构化数据。

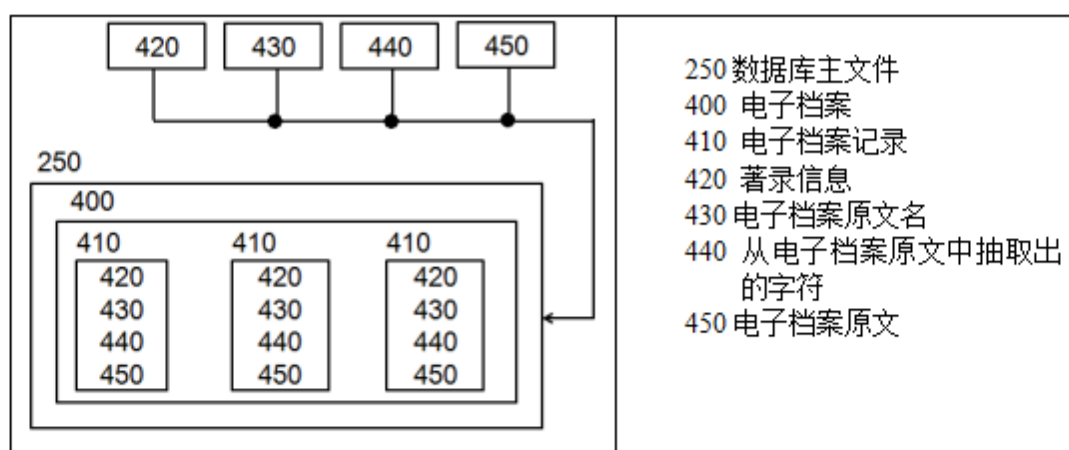


图 9 电子档案录入示意图

3.7.2 电子档案导出

单个或批量导出电子档案（见图 10）。

著录信息和电子档案原文名导出到同一个 csv 或 excel 文件，电子档案原文以原格式导出到文件夹。

存放著录信息和电子档案原文名的 csv 或 excel 文件名和存放电子档案原文的文件夹名相同。

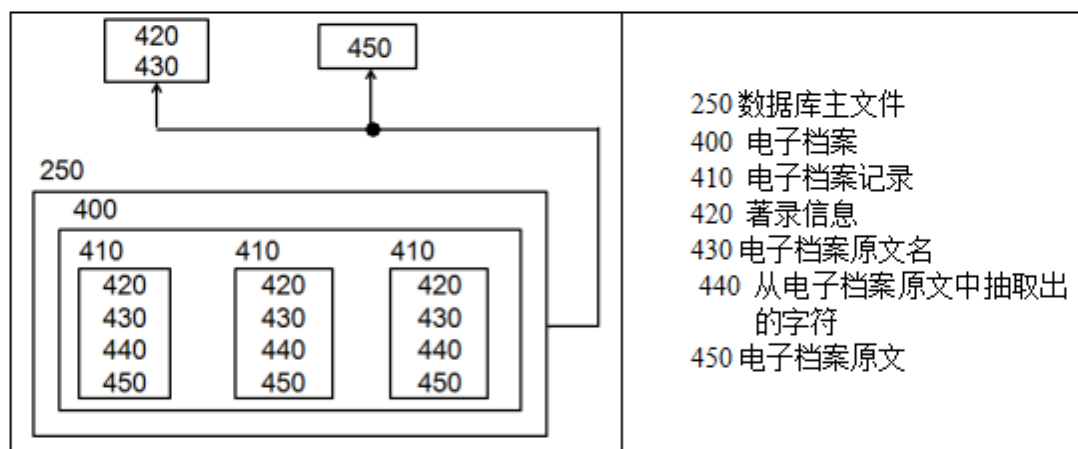


图 10 电子档案导出示意图

4 数据库文件组织

4.1 数据库主文件

数据库主文件命名规则见附录 A。

数据库主文件存储的数据类型包括非结构化数据、半结构化数据、结构化数据。

数据库主文件的数量为一个或多个。

存储数据库主文件的存储介质为磁盘或光盘。

4.2 数据库索引文件

数据库索引文件命名规则见附录 A。

数据库索引文件的数量与数据库主文件的数量相同，一个数据库索引文件对应一个数据库主文件。

存储数据库索引文件的介质为磁盘或光盘。

5 数据库结构设计

5.1 设计原则

数据库结构设计应符合下列原则：

- 数据库结构应满足电子档案存储和应用的需要；
- 数据库结构应按数据分类进行设计；
- 数据模型应采用非关系模型，按非关系数据库的规范化方法进行逻辑设计；
- 数据库名和字段名应命名准确且无二义性，组词和省略应符合习惯；
- 一条记录应可存储非结构化数据、半结构化数据和结构化数据。

5.2 数据库的设计与定义

数据库设计应包括数据库名、数据库存储内容、数据库描述及字段定义设计；数据库名应具有唯一性，体现数据存储内容及其特点。

数据库由记录组成，记录由字段组成，字段类型共有 7 种，包括词组、整数、数值、日期、时间、文本、二进制型字段，其中词组、整数、数值、日期、时间型字段由子字段组成，子字段的数量无限制；文本型字段下分段落，段落下分句子，句子下分词，段落、句子的数量、词的数量无限制。词组、整数、数值、日期、时间型字段用于存储结构化数据；文本、二进制型字段用于存储非结构化数据。

非结构化数据存储于数据库，一般包括三种数据类型的字段：词组、文本、二进制型字段。词组型字段用于存储电子档案名；文本型字段用于存储从电子档案中抽取出来的字符；二进制型字段用于存储电子档案原文。

5.3 字段的设计与定义

字段名应具有唯一性，字段的数据类型和长度应满足存储全部数据的要求。在数据库结构描述中，字段的数据类型应符合以下规定。

——词组型字段以固定长度的空间存储字符数不超过 256 个文字型数据表示，词组型的描述格式为 **Phrase**，缩写为 **PH**；

——文本型字段以可变长度的空间存储文字型数据表示，按段落、句子、词格式编写的自由文体，字符数没有限制。文本型字段可用于存储文本数据、从电子文件中抽取出来的字符。文本型的描述格式为 **TExt**，缩写为 **TE**；

——整数型字段用于存储整数，整数范围为 -2,147,483,647 至 +2,147,483,647。整数型的描述格式为 **INteger**，缩写为 **IN**；

——数值型字段用于存储整数和实数，数值范围为 -1.7E+37 至 1.7E+37。数值型的描述格式为 **NUmber**，缩写为 **NU**；

——日期型字段用于存储公元纪年，日期型的描述格式为 **DAte**，缩写为 **DA**；

——时间型字段用于存储北京时间，时间型的描述格式为 **TIme**，缩写为 **TI**；

——二进制型字段用于存储非结构化数据，包括字处理文件、图像文件、图形文件、音视频文件、多媒体文件、数据库文件、计算机程序、数据文件、超文本文件等，二进制型的描述格式为 **STring**，缩写为 **ST**。

6 存储介质技术要求

6.1 磁盘

参见档案行业对磁盘的相关要求。

6.2 光盘

光盘应使用 **BDR** 档案级蓝光盘，光盘容量为 25GB 及以上，光学性能等技术指标应达到档案行业相关标准要求。

7 光盘标识和光盘匣标识

7.1 光盘标识

在光盘表面黑白打印或向光盘厂定制，包括光盘号、内容摘要、光盘制作时间等内容。

7.1.1 光盘号

由数据库名、档案门类代码、全宗号组成，中间用分隔符连接，表示方法如下：

JYB_001—WS—Z109

其中，数据库命名规则见附录 A；档案门类代码标明光盘内存储档案信息的类型，各档案馆根据各自馆藏情况进行编制；全宗号是国家综合档案馆为同级立

档单位给定的全宗的规范化代码，由 4 位代码构成。

一张光盘存储一个以上的数据库时，数据库名之间用位于中间的“.”相连。

7.1.2 内容摘要

用全宗名称和档号对档案信息内容进行简要说明。

7.1.3 光盘制作时间

光盘制作时间的格式为 YYYY-MM-DD，YYYY、MM、DD 分别表示年、月、日，位数不足时在高位补 0。

7.2 光盘匣命名与标识

光盘匣号由英文字母 M 和数字组成，数字为四位，不足四位填零占位。数据库名规定参见附录 A。

采用条形码、或二维码或 RFID，由光盘匣号、数据库名、光盘卷标（光盘名）组成，中间用分隔符连接，表示方法如下：

M0001—JYB_001—JYB001

一个光盘匣存放一个以上的光盘，表示方法如下：

M0001—JYB_001—JYB001

M0001—KYB_010—KYB010

.....

8 光盘制作流程

8.1 在磁盘上建立数据库的光盘制作流程

8.1.1 建立数据库

在磁盘上建立数据库，然后将数据库转移到光盘上。当数据库容量大于光盘容量时，将该数据库拆分为若干个子数据库，每个子数据库的容量小于光盘容量。

8.1.2 光盘命名

给光盘命名（光盘名亦称光盘卷标名），光盘名与数据库名（数据库容量小于光盘容量）或子数据库名应相同。

8.1.3 打印光盘标识

在光盘盘面上打印光盘标识。

8.1.4 刻录光盘

用单个刻录机刻录时，将光盘放入光驱中刻录数据库；用光盘库刻录时，将

光盘放入光盘库中的光盘匣。光盘库管理系统管理光盘库的所有光盘匣中的全部光盘，文件系统中的一个文件夹代表一张光盘，光盘名（光盘卷标）与文件夹名相同。刻录时只需将数据库（或子数据库）拖入相应的文件夹。

8.2 直接在光盘上建立数据库的制作流程

8.2.1 光盘命名

给光盘命名（光盘名亦称光盘卷标名），光盘名与数据库名相同。

8.2.2 打印光盘标识

在光盘盘面上打印光盘标识。

8.2.3 建立数据库

用单个刻录机时，将光盘放入光驱中，使用非关系数据库管理系统建立数据库。

用光盘库时，将光盘放入光盘库中的光盘匣，光盘库管理系统管理光盘库中的所有光盘匣中的全部光盘，文件系统中的一个文件夹代表一张光盘，光盘名（光盘卷标）应与文件夹名相同，使用非关系数据库管理系统建立数据库。

附录 A
(规范性附录)
数据库命名规则

A.1 数据库命名

数据库命名格式：xxxxxxx；

数据库名由英文字母、数字和下划线组成，须由英文字母开头，最长 16 个字符；

表示方法如下：JYB_001。

A.2 数据库主文件命名

数据库主文件名由数据库名和扩展名组成；

数据库主文件命名格式：xxxxxxx.BAF；

BAF — 数据库主文件的扩展名；

数据库名同 A.1；

表示方法：JYB_001.BAF。

A.3 数据库索引文件命名

数据库索引文件名由数据库名和扩展名组成。

数据库索引文件命名格式：xxxxxxx.BIF；

BIF — 数据库索引文件的扩展名；

数据库名同 A.1；

表示方法：JYB_001.BIF。