



国家图书馆文津出版基金资助项目

家谱谱系

数据模型研究

肖禹 王昭◎著



國家圖書館出版社
National Library of China Publishing House

国家社会科学基金重大项目“中国古籍数字化工程研究”
(项目编号: 12@ZH016) 研究成果

家谱谱系数据模型研究

肖禹 王昭 著



國家圖書館出版社
National Library of China Publishing House

图书在版编目 (CIP) 数据

家谱谱系数据模型研究 / 肖禹, 王昭著. -- 北京 : 国家图书馆出版社, 2016.11
ISBN 978-7-5013-5980-6

I . ①家… II . ①肖… ②王… III . ①家谱—数据模型—研究—中国 IV . ① K820.9

中国版本图书馆 CIP 数据核字 (2016) 第 258589 号

书 名 家谱谱系数据模型研究
著 者 肖 禹 王 昭 著
责任编辑 张 颀

出 版 国家图书馆出版社 (100034 北京市西城区文津街 7 号)
(原书目文献出版社 北京图书馆出版社)

发 行 010-66114536 66126153 66151313 66175620
66121706 (传真) 66126156 (门市部)

E-Mail nlcpress@nlc.cn (邮购)

Website www.nlcpress.com → 投稿中心

经 销 新华书店

印 装 北京玥实印刷有限公司

版 次 2016 年 11 月第 1 版 2016 年 11 月第 1 次印刷

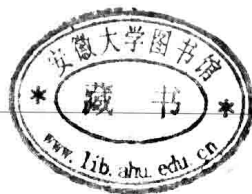
开 本 787×1092 (毫米) 1/16

印 张 20.5

字 数 500 千字

书 号 ISBN 978-7-5013-5980-6

定 价 60.00 元



前言

目前,与家谱相关的专著主要有谱牒学论著、家谱目录提要、家谱资料汇编、家谱专题内容研究等,与家谱数字化相关的专著仍以元数据研究为主。本书以家谱谱系为研究对象,借鉴现有的谱牒学和古籍数字化领域研究成果,对国内外具有代表性的家谱数据模型进行比较研究,并在此基础上研制新的谱系数据模型。

家谱谱系数据广泛应用于寻根 SNS、家谱编撰、文献整理、历史研究等多个领域,在各种类型的谱系中最具代表性。通过对家谱谱系数据模型的研究,可以解决家谱谱系的描述与应用的问题,也可以用于家谱谱系数据交换格式或长期保存。家谱谱系数据模型除了支持一般的应用工具,还要支持专业学术工具,注重信息来源、引用情况、版本比较、文字校雠、专家意见等专业信息的记录和存储,提供必要的研究环境支持。家谱谱系是按照世系关系(包括血缘关系、婚姻关系、继承关系等)组织的人物信息集合,本身就具有数字资源整合的性质,如果加入数据链接,引入时间、空间、文献等维度,可以实现更大范围的资源整合。

本书以家谱谱系为研究对象,通过与家谱索引数据、全文文本数据进行比较,分析了谱系数据模型的基本特征和主要用途。通过对谱系电子表格、世系树 XML、GEDCOM 等现有数据模型的比较研究,从家谱谱系的内容特性出发,结合中华寻根网数据建设项目中出现的实际问题,明确家谱谱系数据模型的设计思路。按照核心集、扩展集、辅助集的方式构建家谱谱系数据模型,核心集由人物子集、关系子集、来源子集、注释子集、链接子集和修订子集组成。人物子集描述谱系中的人物信息,关系子集描述谱系中的人物关系,链接子集描述谱系数据的内部链接或外部链接,来源子集描述谱系数据的来源信息,修订子集描述谱系数据的修订信息。人物子集可进行扩展;关系子集支持家庭模式和宗族模式,宗族模式分为 3 个等级:宗族树,由一棵完整的世系树及若干与宗族相关的孤立节点组成,可用于描述房谱、支谱和世代连续的宗谱;统宗树,由多棵世系树或宗族树及若干与宗族相关的孤立节点组成,可用于描述宗谱和通谱;联宗树,由多棵世系树、宗族树或统宗树及若干与宗族相关的孤立节点组成,可用于描述联宗谱(含同姓联宗谱和异姓联宗谱)。

本书着力于家谱谱系数据模型的分析、比较与构建,希望能够提供一个结构化(或半结构化)的可控数据描述方式,使家谱谱系摆脱传统载体和书写形式的限制,依据应用的需求,订制生成一定复杂度的谱系数据,同时兼顾谱系数据的规范性和可复用性;提供灵活的家谱谱系存储方式,模型的 XML 实现或数据库实现都可以用于家谱谱系数据的保存,加入必要的元数据和存储机制,也可用于家谱谱系数据的长期保存;支持数据影射机制,可以与基于其他模型的家谱谱系数据实现无损(或低信息损失)数据交换;实现结构化人物描述和显性人物关系描述,其数据实现支持世系检索、查重、连缀、截取、拼接、重构、解构等应用,既可用于单机软件,也可用于网络平台;除了支持一般的应用工具,还支持专业学术工具,注重信息来源、引用情况、版本比较、文字校雠、专家意见等专业信息的记录和存储,提供必要的研究环境支持;支持家谱人物信息资源整合,若加入数据链接,引入时间、空间、文

献等维度,可实现更大范围的资源整合。

本书为国家社会科学基金重大项目“中国古籍数字化工程研究”(项目编号:12@ZH016)和国家图书馆科研项目“家谱谱系数字化模型研究”(项目编号:NLC-KY-2014-12)的研究成果,获得国家图书馆文津出版基金的资助,特此致谢!在项目研究和成书过程中,国家图书馆陈红彦、申晓娟、龙伟、李志尧、周晨等老师给予了热心的指点,提出了很多中肯的意见和建议;国家图书馆出版社的金丽萍、王炳乾、张颀等老师为本书的出版付出了辛苦的劳动,在此一并表示感谢!

肖 禹

2016年8月

目 录

前 言	(1)
1 绪论	(1)
1.1 谱系	(1)
1.1.1 谱学	(2)
1.1.2 家谱谱系	(3)
1.2 家谱谱系数据	(7)
1.2.1 家谱数字化	(7)
1.2.2 家谱谱系数据研究现状	(7)
1.3 研究定位	(9)
2 谱系电子表格	(11)
2.1 谱系与电子表格	(11)
2.2 谱系电子表格制作	(12)
2.2.1 谱系电子表格设计	(13)
2.2.2 谱系电子表格录入	(14)
2.3 谱系电子表格特性	(15)
3 家谱树	(19)
3.1 家谱树与 XML	(19)
3.2 家谱树标准规范	(31)
3.3 家谱树特性	(39)
3.4 世系 XML 存在的问题	(41)
3.4.1 定义问题	(42)
3.4.2 数据模型问题	(43)
3.4.3 元素属性问题	(43)
3.4.4 XML Schema 问题	(44)
4 GEDCOM 格式	(49)
4.1 GEDCOM 5.5	(49)
4.1.1 数据表示语法	(50)
4.1.2 世系连接语法	(51)
4.2 GEDCOM 6.0	(58)

4.3	GEDCOM X	(67)
4.4	GEDCOM 数据模型分析	(76)
4.4.1	GEDCOM 5.5 应用	(76)
4.4.2	GEDCOM 到 GEDCOM XML	(102)
4.4.3	GEDCOM X 的启示	(106)
5	世系树	(107)
5.1	待解决的问题	(107)
5.1.1	谱系数据来源的多样性	(107)
5.1.2	谱系的复杂性	(109)
5.1.3	谱系数据应用的差异性	(110)
5.2	世系树模型结构	(112)
5.2.1	世系树核心集	(113)
5.2.2	世系树扩展集	(123)
5.2.3	世系树辅助集	(123)
5.3	世系树模型的 XML 描述	(123)
5.3.1	描述说明	(123)
5.3.2	XML Schema 描述	(138)
5.4	应用实例	(167)
	参考文献	(221)
	附录 1 谱系全文文本数据样例	(224)
	附录 2 谱系索引数据样例 (仍以图 1 为例)	(235)
	附录 3 《福州通贤龔氏支譜》谱系图及世系相关文字摘录 (始祖至十五世)	(237)
	附录 4 世系 XML Schema	(243)
	附录 5 世系连接 GEDCOM 标签 (GEDCOM5.5) 表	(246)
	附录 6 世系连接 GEDCOM 结构和子结构 (GEDCOM5.5) 表	(250)
	附录 7 GEDCOM XML 元素表	(262)
	附录 8 GEDCOM 5.5 标签与 GEDCOM XML 对应关系表	(265)
	附录 9 GEDCOM XML DTD	(270)
	附录 10 家谱元数据 XML Schema	(276)
	附录 11 家谱数据交换标准 XML Schema	(299)

1 绪论

家谱,又称为族谱、宗谱、世谱、谱牒、家乘等,徐建华在《中国的家谱》中列举与家谱相关的名称 70 余个^①,王鹤鸣在《中国家谱通论》中列举有代表性的家谱定义 10 多种^②,欧阳宗书在《中国家谱》中将家谱定义为三个层面:家谱是记述氏族或宗族世系的书;家谱是记录宗族世系和事迹的书;家谱就是家族史^③。本书引用《中国家谱总目·前言》中的家谱定义,“家谱,或曰族谱、宗谱,是记载同宗共祖的血亲集团世系、人物、事迹的历史图籍”^④。

1.1 谱系

谱系,在《现代汉语词典》上有两个定义:①家谱上的系统;②泛指事物发展变化的系统^⑤。

第一个定义可以理解为狭义的谱系,家谱中的人物与人物关系所构成的系统,即世系。欧阳宗书在《中国家谱》中指出:“世系又称世传录、世系表、世系图等,它是家谱中的主体部分。这部分内容是中国历代家谱所共有的,家谱之所以能区别于正史与方志,主要就由这部分内容来体现。它详细记载从一世祖至修谱日期止,家族所有成员的姓名、生卒年月日、寿数、科第、职官、葬地、妻室、正妻及续娶的姓氏、父名、行次、有子女妾姓氏、受封侧室姓氏。子,人数,有无出嗣。女,人数,出嫁何人。这些一律依辈分高低、年龄长幼、嫡庶亲疏排列。一族人的源流关系、血缘亲疏、繁衍情况、婚嫁情况皆可一目了然。”^⑥徐建华在《中国的家谱》中认为:“世系,也称世表、世系表、世系图、根图,是以图表形式反映家族成员的血缘关系,这是家谱的主要内容,通常是五代为一表。世系录,也有家谱称世序、世系考、传实、行实、世录,是对世系表的解释,即记录一个人生、老、病、死、葬的简历,内容包括父名、排行、名、字、号、生卒年月日时、享年、官职、功名、德行、葬地、葬向、妻妾的生卒年月日时、封诰、岳家、子女、女嫁之人,有无富贵外孙等,特别重生死、血统。”^⑦王鹤鸣在《中国家谱通论》中将世系定义为“记载一世祖至撰修该家谱时历代成员的姓名字号、生卒年月、职官科第、妻室子女、葬地坟茔等内容的图表。它是中国家谱的主体部分和后人谒祖寻

① 徐建华.中国的家谱[M].天津:百花文艺出版社,2002:9.

② 王鹤鸣.中国家谱通论[M].上海:上海古籍出版社,2010:1-3.

③ 欧阳宗书.中国家谱[M].北京:新华出版社,1993:1-3.

④ 王鹤鸣.中国家谱总目[M].上海:上海古籍出版社,2008:前言.

⑤ 中国社会科学院语言研究所词典编辑室.现代汉语词典[M].修订本.北京:商务印书馆,2000:989.

⑥ 欧阳宗书.中国家谱[M].北京:新华出版社,1993:114-115.

⑦ 徐建华.中国的家谱[M].天津:百花文艺出版社,2002:37-38.

根的主要线索,也是区别家谱与其他方志、家传、家史等书籍的主要依据。世系居家谱中重要分量,往往占到一部家谱四分之三以上的篇幅”^①。

第二个定义可以理解为广义的谱系,泛指事物发展变化的系统,除了家谱谱系以外,还可以是植物谱系、动物谱系、器物谱系、神仙谱系、学术谱系等。以印度因明谱系为例,如图1-1所示,学术谱系中包含代表人物、师承关系、学术观点等内容。再以《真灵位业图》的神仙谱系为例,该书将之前道书中所出现的先天尊神、后天仙真、人鬼、星君、地祇、幽冥官吏及各类神将、仙官等近七百位神鬼仙圣,纳入其所设定的玉清、上清、太极、太清、九宫、洞天、酆都等七个等级的体系之中,其中每一级都有一位居于中位的主神统领,左右分设众多辅神拱卫守护,形成了一个等级森严、和谐有序的神仙谱系^②。

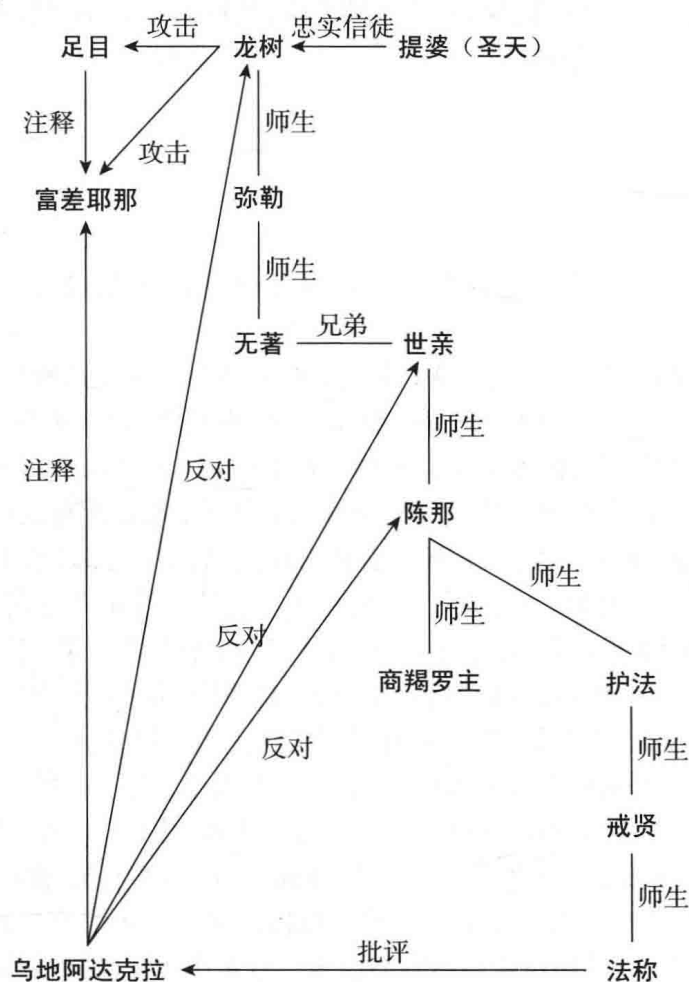


图 1-1 印度因明谱系图^③

1.1.1 谱学

谱学,也称作谱牒学、家谱学等,即谱牒之学。白寿彝在《中国史学史》中指出:“魏晋

① 王鹤鸣.中国家谱通论[M].上海:上海古籍出版社,2010:306.

② 李剑楠.道教神仙谱系的建构初探[D].北京:中央民族大学,2010:40.

③ 刘纯.由因明谱系看其思想演变[D].秦皇岛:燕山大学,2009:9.

南北朝隋唐时期历史撰述中反映门阀地主的要求和趣味,就颇为显著。其中之一就是由名门大姓家谱之分别编撰逐渐发展成为一门综合各家记载的专门之学——谱学。”^①仓修良在《试论谱学的发展及其文献价值》中认为:“谱学和方志学一样,都是史学的旁支,并随着史学的发展而产生和形成,乃至最后成为一门独立的学问——谱学。”^②欧阳宗书在《中国家谱》中将谱学定义为:“专门研究谱牒,亦即家谱的学科。其内涵就是:研究中国家谱产生、发展的过程及其赖以生存的条件,揭示其发展规律,并科学地探究家谱所蕴含的各种文化内容和它的意义。谱牒学的外延应该是:从历史学、宗教学、法学、教育学、经济学、伦理学、人才学、遗传学等众多学科的视角出发,对中国国家谱作综合地、立体地、全面地研究,以揭示家谱的特有历史功能、学术价值及现实意义。谱牒作为一门学问,虽然早在魏晋南北朝时期即已出现,但过去的所谓‘谱学’,无非是讲如何修家谱,如何更得力地、更密切地为特权阶级服务,远非现实意义上的、具有完整理论体系的独立学科。”^③

“谱学”一词的出现不应晚于南北朝时期,在《南齐书·贾渊传》中有“贾渊,字希镜,平阳襄陵人也。祖弼之,晋员外郎。父匪之,骠骑参军。世传谱学。……先是谱学未有名,渊祖弼之广集百氏谱记,专心治业”^④。王鹤鸣在《中国家谱通论》中指出,《世本》是周代谱学著作的代表,也是后代所有总括性的谱牒所宗师、祖述的开山之作,《世本》等谱系著作的问世,表明中国谱学已脱离原始形态,标志着中国谱学已正式诞生^⑤。仓修良在《试论谱学的发展及其文献价值》中概述了中国谱学发展的三个高潮:魏晋南北朝时期,门阀豪族最重门第、血统、婚宦,谱学成为维护门阀豪族利益、巩固门第制度的一种史学,修撰家谱的风气盛极一时,出现了许多著名的谱学家和专著,形成了谱学发展的第一个高潮;到了唐代,谱学在魏晋南北朝的基础上继续得到发展,为了提高政治集团的社会地位、调节地主阶级各阶层之间的关系、巩固政权统治,唐代的谱学主要掌握在政府手中,几部大的谱牒著作亦皆为官修,出现了谱学发展史上第二个高潮;到了唐代后期,由于士、庶界限逐渐淡漠,矛盾、斗争逐渐缓和,官修谱学也就随之衰落下去,宋代开始,大多以一家一族修“私家之谱”为主,以欧、苏两家最为出名,后世奉为法式,同时,产生了一种新的体裁——年谱(按年月专门记载某一个人生平事迹的一种著作,它是由传记体发展而来的),特别到了清代,年谱之风盛行,乾嘉时期风气尤盛,遂出现了谱学发展史上第三个高潮^⑥。

1.1.2 家谱谱系

家谱谱系,即世系,是家谱的核心内容。不论哪一类型的谱书,均有宗族得姓渊源、世系、地望及代表人物小传,且以世系为主体,若无世系就不成其为宗族谱了^⑦。目前已知最早记录世系文字的是三件甲骨:《库方二氏藏甲骨卜辞》之1506号,容庚、瞿润缙撰《殷契卜辞》

① 白寿彝.中国史学史:第一册[M].上海:上海人民出版社,1989:58.

② 仓修良.试论谱学的发展及其文献价值[J].文献,1983(2):207.

③ 欧阳宗书.中国家谱[M].北京:新华出版社,1993:136.

④ 萧子显.南齐书[M].北京:中华书局,1972:906-907.

⑤ 王鹤鸣.中国家谱通论[M].上海:上海古籍出版社,2010:62-63.

⑥ 仓修良.试论谱学的发展及其文献价值[J].文献,1983(2):211-229.

⑦ 冯尔康.中国古代的宗族和祠堂[M].北京:商务印书馆,1996:167.

之 209 号和董作宾编《殷虚文字乙编》之 4856 号^①。现存最早的纸本世系是《高昌某氏残谱》(1973 年出土于阿斯塔那 173 号墓,收入《吐鲁番出土文书》第三册第 64 页)和《某氏族谱》(1966 年出土于阿斯塔那 50 号墓,收入《吐鲁番出土文书》第三册第 179 至 184 页)^②。

钱杭在《宗族的世系学研究》中指出,“世系的基本特征,是通过排列有序、世次分明的文、图、表等形式,连续性地描述人类上下世代之间的继承关系^③”,并列出了 7 种典型的世系:无配偶、无旁系、无世序、无间断;无配偶、无旁系、有世序、无间断;无配偶、有旁系、无世序、无间断;无配偶、有旁系、无世序、有间断;无配偶、有旁系、有世序、有间断;有配偶、无旁系、无世序、无间断;有配偶、有旁系、有世序、无间断。

倪义省在《家谱编撰概说》中总结了历代世系编撰方式的变化:汉代以前,一般只按照世系先后,简略叙述家族的人名,交代人物之间的世系关系,形式非常简单;汉代,世系不仅记人名(即世谥)、字号,而且记官位、姓氏起源等,形式上有横格制(分代分格按顺序记载)、(以姓为单位)先叙述得姓渊源再叙述世系官位、一以贯之的叙述;魏晋南北朝时期,世系开始反映婚姻情况,出现了现存最早的世系图表;唐代,世系以图表式为主,一般分代分格填写;宋代,欧、苏两种世系编排都行“小宗之法”,五世一图,也有欧、苏并用(世系图用欧式,世系表用苏氏);元代,不拘泥于欧、苏的“小宗之法”,以“大宗之法”绘制谱图,出现了女儿、女婿、继室、侧室、继子等入谱的情况;明、清、民国,世系编撰方法在欧、苏两式的基础上不断发展,日臻完善^④。

宋代是中国家谱的转型时期,欧阳修和苏洵创立了新的谱法,对后世影响巨大,即“欧苏体例”。欧式谱系按世代分格,自上而下每格记一世,五世为一图(五世以后,格尽另起),谱名左侧小字记字号、生卒、配偶、目的、职官等,如图 1-2 所示。与欧式相比,苏式更强调宗法关系。苏式谱系仍然五世为一图,世代直行下垂,世代间用竖线连接,同辈自右向左(由大到小)排列,如图 1-3 所示,五世之后另起一图,但是第五世必须要同见于第一图末和第二图始,因此,第二图记第五至九世,第三图记第九世至十三世,依次类推。苏洵谱法一方面强调从纵的方面,注意诸父多子相继关系,主张五世则迁的小宗之法;另一方面又强调从横的方面,对兄弟分支加以却别,推崇合各支为一编的大宗谱,推崇大宗之法^⑤。后世修谱亦有欧苏并用的形式,即世系图用苏式,世系表用欧氏,如图 1-4 所示。

王鹤鸣在《中国家谱通论》中分析了徽州(一府六县)明至民国间 48 部家谱,世系都遵循欧苏体例,可以细分为 4 种情况:五世一图,第五世另起一图为图首,世系与世传合图;五世一图,第六世另起一图为图首,世系与世传合图;五世一图,第五世另起一图为图首,世系与世传分列;五世一图,第六世另起一图为图首,世系与世传分列。又任选 50 种徽州家谱,第 1 种情况占 30%,第 2 种情况占 20%,第 3 种情况占 30%,第 4 种情况占 10%,其他情况占 10%^⑥。

除了欧苏体例以外,家谱谱系还有宝塔式、牒记式、韵文式等。

① 陈梦家.殷虚卜辞综述[M].北京:中华书局,1988:499.

② 郭锋.晋唐时期的谱牒修撰[J].中国社会经济史研究,1995(1):29.

③ 钱杭.宗族的世系学研究[M].上海:复旦大学出版社,2011:43.

④ 倪义省.家谱编撰概说[M].福州:福建人民出版社,2009:117-123.

⑤ 王鹤鸣.中国家谱通论[M].上海:上海古籍出版社,2010:114.

⑥ 王鹤鸣.中国家谱通论[M].上海:上海古籍出版社,2010:403-407.

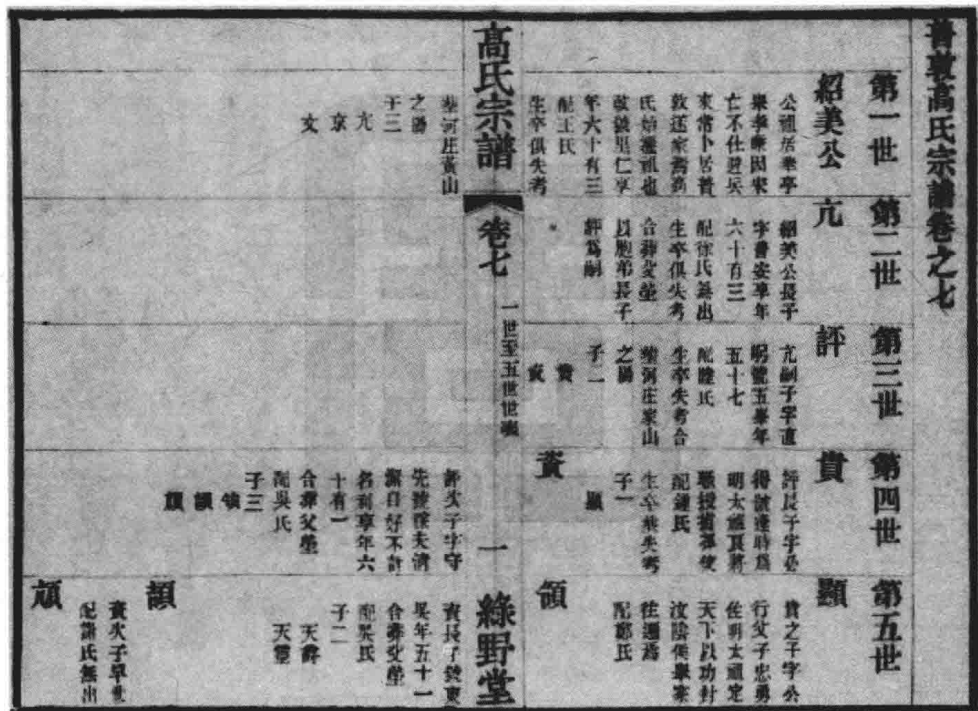


图 1-2 欧式谱系样例图①

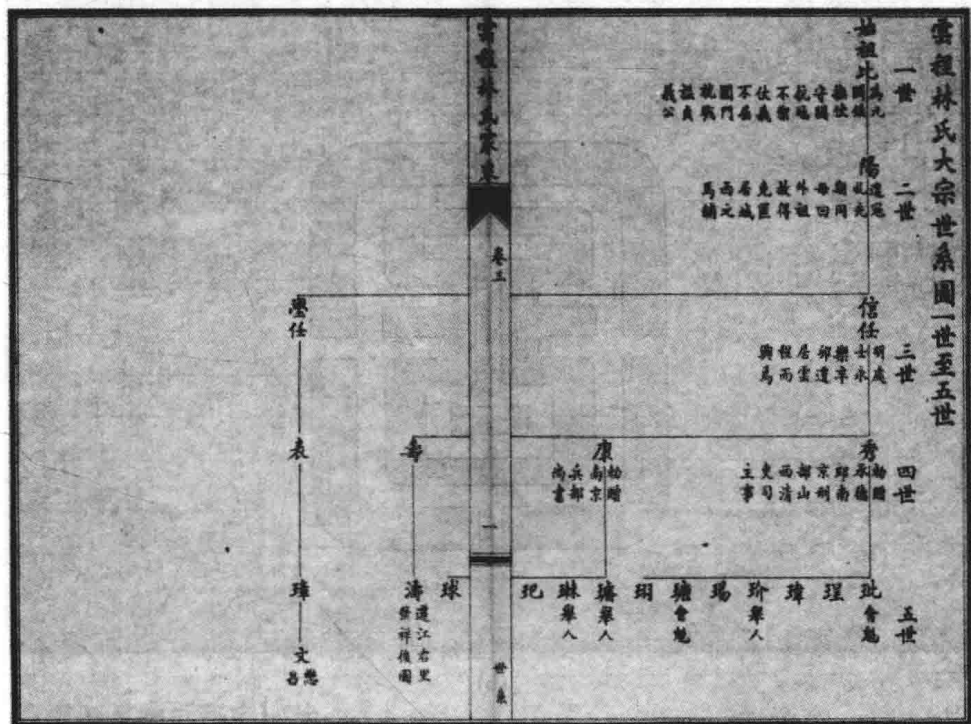
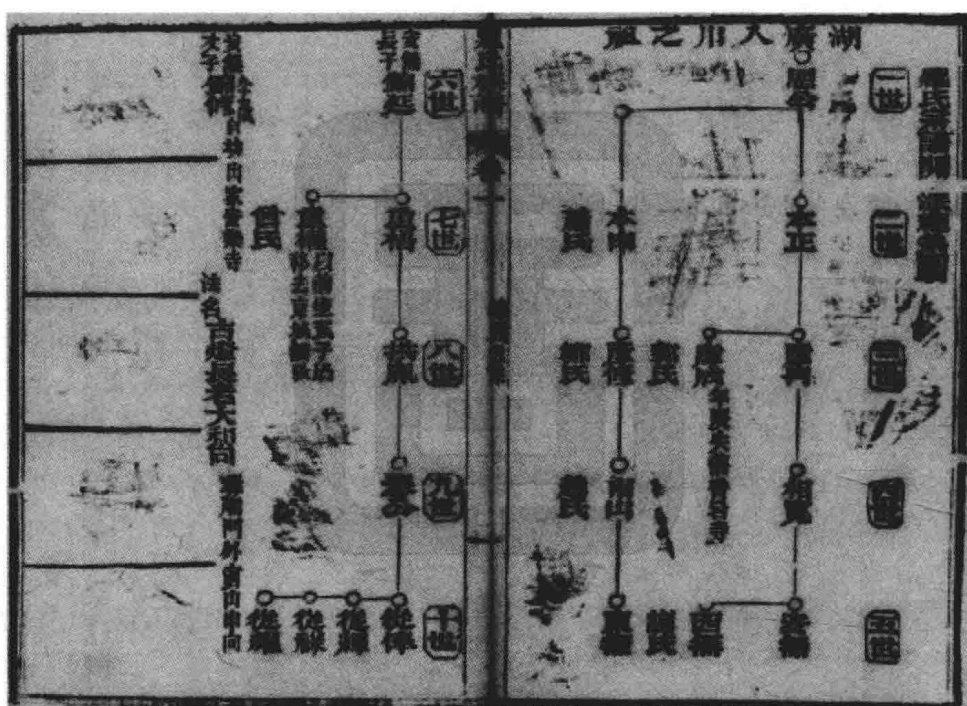


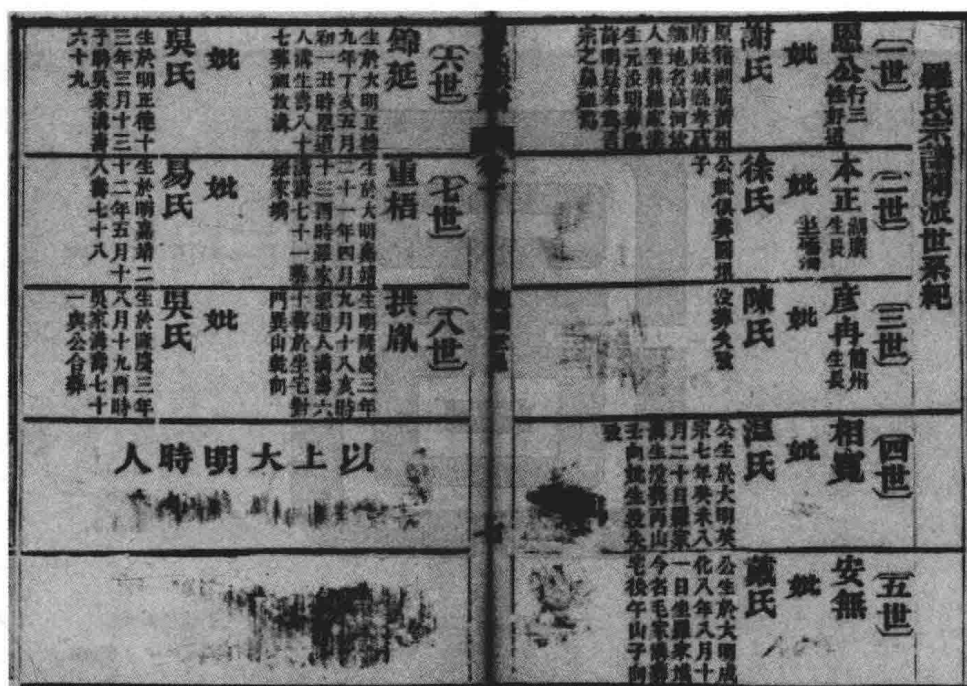
图 1-3 苏式谱系样例图②

① 见：高近纲等纂修《普敦高氏续修宗谱》(民国二十年绿野堂木活字本)。

② 见：林懋勛等纂修《侯官云程林氏家乘》(民国二十七年铅印本)。



(a)



(b)

图 1-4 欧苏并用谱系样例图^①

① 见：罗元廷等纂修《罗氏族谱》(民国二十年增修刻本)。

1.2 家谱谱系数据

传统的家谱谱系载体有甲骨、青铜器、石、绢布、纸等,现代的载体有缩微胶片、磁盘、光盘等。将家谱谱系数字化或在数字环境下编制谱系,即可生成家谱谱系数据。家谱谱系数据的存储格式有 TXT、DOC (或 DOCX)、HTML、XML、TIFF、JPEG、PDF、XLS (或 XLSX)、GED、MDB (MACCDB)、MYD、MDF 等。除了口传或结绳谱系采用音视频方式保存,使用文字编制的谱系通常采用图像或文本方式保存。

1.2.1 家谱数字化

谱系是家谱的核心内容,且占据家谱文献的主要篇幅。因此,可以通过考察家谱数字化情况了解家谱谱系数据的基本情况。毛建军在《中国家谱数字化资源的开发与建设》^①中概述了国内家谱目录数据库和全文数据库建设的基本情况,在《中国家谱数字化的新进展》^②中概述了“家谱先生”“天下家谱”等家谱编辑软件和“中华寻根网”等家谱数字资源整合系统。王昭在《家谱文献资源数字化现状与思考》^③中概述了国内家谱文献数字化建设现状,图书馆、博物馆等公藏机构和商业机构、宗亲会等已取得的成果,并就家谱数字化工作分散进行、无统一标准、重复建设、无法整合、使用不便等问题提出对策。

目前,家谱数字化的主要形式有目录、索引、图像、文本等。目录数据主要描述家谱文献的外部特征,揭示家谱文献的内容特征,不能直接用于描述家谱谱系。图像数据只能反映谱系在家谱文献中的原貌,不能支持检索等应用。全文文本数据描述了家谱文献的内容信息和版式信息,可以支持全文检索,如附录 1 所示。page 标签描述叶(筒子页)信息,首先是叶号、对应的图像文件、图像大小、图像中心等,然后用 text_region_list 标签描述文字和文字版式,用 page_text_list 描述版心,用 image_list 描述叶内插图,用 line_list、rect_list 等描述叶内的图形,reverse_list 描述集外字贴图。通过对附录 1 的分析可以发现全文文本数据将表示世系关系的线段与版框、界栏等都作为图形处理,并没有语义标记,既没有检索意义,也无法用于显示以外的应用。如果只从检索角度考虑,全文文本数据与附录 2 中的索引数据差别不大。当然,附录 2 中的索引数据并不是典型意义上的索引数据,传统的人名索引数据要去掉 generation、branch 和 detail 标签。

家谱谱系数字化当然属于家谱数字化范畴,但是由于谱系格式与内容的复杂性,传统的目录、索引、图像、文本等数据都不能满足谱系数字化的要求,谱系数据应该是结构化人物描述,显性人物关系描述,支持世系检索、查重、连缀、截取、拼接、重构、解构等。

1.2.2 家谱谱系数据研究现状

目前,国内对家谱谱系数据的研究还比较少,择其要者简述如下:

① 毛建军. 中国家谱数字化资源的开发与建设[J]. 档案与建设, 2007(1): 22-24.

② 毛建军. 中国家谱数字化的新进展[J]. 数字与缩微影像, 2013(2): 47-48.

③ 王昭. 家谱文献资源数字化现状与思考[J]. 科技情报开发与经济, 2013(10): 110-113.

张帆在《家谱档案转型研究》^①中指出,国内外现有电子家谱档案有两种制作方式,一是通过的大型综合性网站中的“家族树”板块或者专门的家族树网站,自行创建电子家谱。二是家谱软件制作电子家谱。家族树通常以一个人作为节点,链接家人,还可以邀请其他亲戚加入,建立家族圈,依托树状结构将家庭中的成员网聚在一起,用户自动生成一个免费的电子家谱,国内网站有:家族网(Jiazuwang.net)、家族(Jia.qidian.com)、家族树(Jiazushu.net)等,国外的有My family.com、Ancestry.com等。通过这些网站编撰的电子家谱,约占所有电子家谱的95%。电子家谱编撰软件与网站家族树模式类似,创建方式和步骤也类似,不同之处在于多为单机版,并且增加了后期的打印功能,国内软件有:家谱先生、弘网家谱、千秋家谱、中华族谱、永恒族谱等,国外有:Agelong Tree、Osk、GenSmarts、Genbox Family History等。目前使用软件编撰的家谱不多,仅占有所有电子家谱的5%。

皇甫冉在《多媒体族谱数据库及网上查询系统》^②中提出,将族谱资料分解为三个相关实体:族谱基本信息、族谱世系信息和族谱补充资料。族谱基本信息包括姓氏编号、族谱序号、族谱名、省、市、乡镇、村、乡镇/村的别名、开基祖先、发源地、起源时间、族源信息(来源及繁衍过程)、衍派(灯号)。族谱世系信息包括世系序号、族谱序号、世代、姓名、字、号、生日、卒日、旅居地(国家)、亡地、葬地、妻子、子女、父亲、父指针。族谱补充资料包括族谱序号、世系序号、资料序号、其他文字资料、多媒体信息说明、多媒体信息(图像或视频资料)。

冯一茹在《家谱元数据规范设计与XML实现》^③中提出,家谱元数据规范包括4个核心实体集和3个辅助实体集,核心实体集包括头信息实体集、书目管理信息实体集、内容结构信息实体集和宗族信息实体集,辅助实体集包括时间信息实体集、空间信息实体集和个人/单位信息实体集。内容结构信息包括:谱序、谱例、谱论、像赞、世系、传记、艺文著述、其他8个子集,目录、修谱名目、姓氏源流、恩荣录、字辈、仕宦录、礼仪、祠堂、墓茔、族规家训、契约、族产、领谱字号、余庆录14个元素。宗族信息包括姓氏、郡望、堂号、房系、支系、备注6个元素,世代范围、时间范围、修谱历史、谱籍、始祖、名人6个子集。详见附录10。

柯洲在《数字化家谱的研究与实现》^④中提出,家谱数据库由家谱总目表、成员信息表和家谱信息表组成。家谱总目表包括家谱ID、家谱名、家谱人数、家谱属地和家谱创制时间字段;成员信息表包括成员编号、家谱编号、姓、名、性别和成员生日字段;家谱信息表包括家谱编号、家族故事、家族历史和家族今日字段。采用静态树状结构表示人物关系。

雷晖在《基于Flex与J2EE的族谱网的设计与实现》^⑤中提出,族谱树采用树的思想,把每个成员的关系以XML的形式存储于数据库中,族谱树以Flash形式展示给用户,Flash根据由成员关系生成的XML文件来生成一棵树的形式,生成XML与提取这部分功能是用PHP来实现,生成的XML文件被保存在数据库中。“族谱树”以我为中心,可以添加父亲、母亲、兄弟、姐妹、妻子和儿女。

① 张帆. 家谱档案转型研究[D]. 天津: 天津师范大学, 2012.

② 皇甫冉. 多媒体族谱数据库及网上查询系统[D]. 泉州: 华侨大学, 2001.

③ 冯一茹. 家谱元数据规范设计与XML实现[D]. 南京: 南京师范大学, 2011.

④ 柯洲. 数字化家谱的研究与实现[D]. 武汉: 华中师范大学, 2011.

⑤ 雷晖. 基于Flex与J2EE的族谱网的设计与实现[D]. 苏州: 苏州大学, 2011.

杨翼在《家谱数据交换标准研究》^①中提出,家谱数据模型由家谱元数据信息、家谱基本信息和家谱时空信息三部分组成。家谱基本信息包含个人信息、家庭信息、家族信息和家谱事件信息。家谱数据交换总体框架由4个主要元素集合和4个次要元素集合组成(次要元素集合不能独立存在)。主要元素集合包括元数据信息、个人信息、家庭信息和家族信息;次要元素集合包括事件信息、文献信息、多媒体信息和引用信息。详见附录11。

国外对家谱谱系数据的研究比较早,20世纪80年代就制订了GEDCOM(Genealogical Data Communication)标准,并广泛应用于家谱软件 and 家谱网站中,类似的标准还有GREnDL(Genealogical Record Exchange and Description Language)、GedXML(the XML specification of GEDCOM)、GenTech Genealogical Data Model、GenXML等。

1.3 研究定位

本书定位于家谱谱系的数据模型研究,家谱谱系广泛应用于寻根SNS(Social Networking Services,社交网络)、家谱编撰、文献整理、历史研究等多个领域,在各种类型的谱系中最具代表性;给家谱谱系建立数据模型,不针对特定软件和应用来设计数据结构,保证数据模型的功能性、通用性和开放性。家谱谱系数据模型可以通过数据文件、数据库等形式实现,也可以形成特定领域内的标准规范。

家谱谱系数据模型具有以下作用:

(1) 家谱谱系数据模型提供了一个结构化(或半结构化)的可控数据描述方式,使家谱谱系摆脱了传统载体和书写形式的限制,依据应用的需求,订制生成一定复杂度的谱系数据,同时兼顾谱系数据的规范性和可复用性。

(2) 家谱谱系数据模型提供了一种家谱谱系存储方式,模型的XML实现或数据库实现都可以用于家谱谱系数据的保存,加入必要的元数据和存储机制,也可用于家谱谱系数据的长期保存。

(3) 一个定义良好的家谱谱系数据模型可以描述家谱谱系中的必要信息,通过一定的数据影射机制,可以与基于其他模型的家谱谱系数据实现无损(或低信息损失)数据交换。

(4) 家谱谱系数据模型实现了结构化人物描述和显性人物关系描述,其数据实现支持世系检索、查重、连缀、截取、拼接、重构、解构等应用,既可用于单机软件,又可用于网络平台。

(5) 家谱谱系数据模型除了支持一般的应用工具,还要支持专业学术工具,注重信息来源、引用情况、版本比较、文字校讎、专家意见等专业信息的记录和存储,提供必要的研究环境支持。

(6) 家谱谱系是按照世系关系(包括血缘关系、婚姻关系、继承关系等)组织的人物信息集合,本身就具有数字资源整合的性质,如果加入数据链接,引入时间、空间、文献等维度,可以实现更大范围的资源整合。

本书第2部分为谱系电子表格,分析利用电子表格存储家谱谱系的利弊,谱系电子表格是一种用户自定义的谱系数据模型;第3部分为家谱树,讨论世系关系的树形结构描

^① 杨翼.家谱数据交换标准研究[D].南京:南京师范大学,2009.

述、人物结构化描述和世系 XML, 世系 XML 是文献收藏机构使用的谱系数据模型; 第 4 部分 GEDCOM, 研究家庭树模型、GEDCOM 文件格式、从 GEDCOM 到 GEDCOM X 的发展变化以及 GEDCOM 如何应用于中国家谱, GEDCOM 是家庭史等专门领域使用的谱系数据模型; 第 5 部分为世系树, 以谱牒学为理论基础, 对上述三类谱系数据模型进行比较研究, 结合家谱数字化项目中出现的实际问题 and 应用需求, 提出世系树模型, 给出世系树文件的 XML schema 描述, 探讨世系树模型的扩展与应用。