

中国肿瘤登记数据管理与分析系统 软件开发与实践

Software Development and Practice of Cancer Registration Data Management and
Analysis System in China//WANG Qing-sheng, CHEN Wan-qing, ZHANG Si-wei

王庆生¹, 陈万青², 张思维²

(1.天津医科大学附属肿瘤医院预防医学中心, 天津 300060; 2.全国肿瘤登记中心, 北京 100021)

摘要:肿瘤登记工作是肿瘤预防和控制的一个重要组成部分,但是目前国内各肿瘤登记处使用的数据登记和管理软件存在较多问题。全国肿瘤登记中心委托天津医科大学附属肿瘤医院研发了一款拥有自主知识产权、简单、易用、可普及的肿瘤登记专业软件。全文对该软件的研发过程、主要特点及应用等方面作一介绍。

关键词:肿瘤登记;软件开发;数据管理

中图分类号:R730.1 **文献标识码:**A **文章编号:**1004-0242(2012)02-0086-06

1 背景

肿瘤登记为肿瘤预防与控制的一个重要组成部分。自20世纪60年代,我国的上海、河南林县等地率先建立肿瘤登记制度。2008年,卫生部设立中央转移地方支付专项经费支持肿瘤登记工作,目前已经有市、县级肿瘤登记处近200个,肿瘤登记覆盖人群达全国人口的13%以上。

国际癌症研究中心(IARC)和国际肿瘤登记协会(IACR)是从事肿瘤登记工作的国际性组织,为其会员提供免费的肿瘤登记应用软件,主要有用于登记数据管理的CanReg 4,以及用于数据转换和审核的IACRtools。目前,全球许多肿瘤登记处都在使用,而且已经升级到CanReg 5。

全国肿瘤防治研究办公室在2002年对CanReg 4软件进行了汉化,并在国内推广至今,目前已经授权近100多个登记处使用。IACRtools没有汉化版,但通过培训,已在各登记处数据处理时广泛使用。此外,北京、上海、江苏、武汉、杭州、广州、大连等登记处都有自己开发或委托软件公司自行开发的肿瘤登记软件。

CanReg 4软件在中文处理、查重、版面设置等方面不太适合中国人习惯,同时开发较早,与目前的计算机软件的便捷性与高效性相距甚远。同时,与其他统计软件和数据库软件数据接口的通用性较差。

在国内肿瘤登记处自行开发的软件中,其中不乏独树一帜的设计理念。但是,这些软件所存在的局限性主要包括以下几点:①没有完整的知识产权,进一步开发和修订受限。②太适应于开发者登记处的本土化,不利全国推广。③公司的开发和修订费用较高。④小型软件因费用问题,计算机专业人员与使用者的沟通和实现存在较大障碍,无法随时发现问题或有新的设计,随时修订和升级。⑤数据库格式不同会加大全国或多中心合作时的工作量。

基于以上所述,如何解决这些问题已经纳入中国肿瘤登记专业人员的工作日程。一个拥有自主知识产权、简单、易用、可普及的肿瘤登记专业软件系统应运而生。

2 软件开发立项、设计理念

在2009年下半年的一次研讨会中,专家们不谋而合地认为开发一款国产化的肿瘤登记软件势在必行。其设计理念就是:开发高度整合的以批处理方式

收稿日期:2011-06-30

通讯作者:王庆生, E-mail: QSWANG@mail.zlnet.com.cn

运行、具有高安全性、易用和可靠的中文版《肿瘤登记数据管理与分析软件》。

其后,全国肿瘤登记中心多次组织专家论证,确定研发方案,并得到卫生部疾病控制局的大力支持,于2010年5月正式立项。在多方评估计算机软件公司和一些单位后,最后委托天津医科大学附属肿瘤医院作为软件主要研发单位。

软件的框架模式:经过多次论证认为对于多数基层单位来讲,单机版比局域网版对计算机技术的要求较低,系统更加稳定,同时比网络版也节省运行时的人力与物力预算。因此,软件框架确定为 Windows 视窗的单机版软件。当然,在有较大需求时,在单机版基础上二次开发为局域网版和互联网为基础的网络版是可行的,也是易行的。

3 软件的主要特点

软件的主要特点就是功能齐全,实用易用。

软件设计使用的编程软件是正版的 Visual Studio 2010 版,数据库管理软件是微软公司最新推出的免费版 SQL Server Express 2008 R2 版。从软件的开发到软件的使用没有知识产权违规的问题。

软件的视窗界面设计合理,数据管理界面为单一界面,输机时不用反复翻页,并可同时浏览相关内容。报卡图像化简化了大量卡片柜的设置和当需要查阅原报卡或相关报卡时翻卡片的烦杂,也利于基层单位用电子文档传送报卡。多窗口信息统一管理和显示便于实际操作、对比信息。为了日常应用的方便,对于病例管理和人口管理设立快捷键,以及窗口快速排列的快捷键。

用户管理、权限设定、用户使用记录查阅等措施更有利于数据的安全。数据备份和可选的多硬盘备份有利于数据的保护。与 Excel、Access 相比,SQL Server 对于数据安全方面更优。

对于基层的肿瘤登记工作人员来说,ICD 编码始终是个较为困难的事情。为了解决此问题,软件在输机界面提供了模糊查询辅助输机模块。当遇到疑难问题时,此模块会提示可能编码供参考。此外,在软件中还设置了专用的 ICD 形态学与解剖学模糊查询页面,用于单独查询和日常复习。模糊查询时,最多可设置 6 个关键词,并且不限前后顺序。

本软件设有导入和导出模块用于兼容其他的数据库和软件。为了迎合现有 CanReg 4 用户,对于

CanReg 4 数据设有快捷导入模板,用于快速转入新系统。

本软件力图囊括较多的数据报表和统计功能。在数据查重、查错、质量控制、发病率、死亡率和其他统计报表等模块将在使用中不断完善和发展。

本软件还设立了专用链接,方便用户打开肿瘤登记相关的网站和有关的子目录,也设置了对于计算器的快捷方式调用。为了各个肿瘤登记处个性化的需求,软件 Logo 界面可以重新设定自己喜欢的图像,重新输入 Logo 界面《中国肿瘤登记中心》下面的一行字,换成自己登记处的名称,并用自己中意的颜色显示。

4 软件框架、数据库与编程语言选择

4.1 软件框架

目前计算机软件主要分为 Windows 视窗软件和网页浏览器的网页软件。前者的优势是数据安全、易于管理、性能更强等。因此本次开发的计算机软件选用前者。框架的编程使用 Visual Studio 2010 的 Visual Basic 语言。

4.2 数据库使用 SQL Server Express R2 版

此种数据库的优势是比 Access、dBase、Foxbase、Epi Info、Excel 等更强,是微软产品。此版本是免费版,最大可管理 10G 大小的数据库,对于多数肿瘤登记处已经够用。如果需要进一步扩大容量,只需购买标准版或更高版本的版权即可直接升级。

4.3 数据处理和运算使用 Fortran 语言编程

Fortran 语言对于数据处理的能力要远优于其他语言,运算速度比数据库程序更快的多。

4.4 系统主视窗界面

开机后首先进入用户登录视窗界面,然后进入主视窗界面,见图 1。主视窗界面包括的主目录有“文件”、“编码”、“数据管理”、“文件管理”、“系统管理”、“窗口”、“帮助”、“窗口速排”、“病例输机”(快捷键)、“人口输机”(快捷键)和“当前用户与相应权限”(仅显示)一栏。

5 数据安全性与用户管理

系统启动后首先打开用户登录窗口,见图 2。登录窗口视窗界面左侧为可以自定义的肿瘤登记处

LOGO,同时有用户名、密码栏、用户分类及相应权限等栏。如果用户三次输入错误密码,系统将强制退出。涉及“用户管理”的数据库包括用户库和“用户使用记录”库。后者记录了登录用户名、登录时间和退出时间。用户分为四类,即为超级管理人员、管理人员、输机人员和使用人员,权限逐级递减。使用人员没有使用病例、人口数据库以及修改自己密码的权限。只有超级管理人员可以设置新用户和浏览用户使用记录,同时也可以看到和修改其他用户的密码。为了管理可以设置一个或一个以上超级管理员用户,并且不要删除最后一个超级管理员用户。

在数据安全方面,只有超级管理人员有权对全部数据库清零、批量导入和导出数据。管理人员可以对数据库备份,并且可以在多个硬盘上分别备份。输机人员可以操作数据库进行分析和统计。使用人员可以浏览基本框架的同时,可以用 ICD 模糊查询功能进行编码查询。

6 数据管理

病例输入可以有两种形式:通过“文本文件导入”的方式导入既往存储的数据库数据。对于 CanReg 4 用户,本软件系统设立有专用的导入模板,可以方便地把过去的 CanReg 4 数据导入新系统。对于其他格式,可以根据诸变量在文本文件中的位置,自行设定导入模板。对于新的报卡,可以通过数码相机或扫描仪把报卡转为图像,也可由报卡单位通过电子邮件、QQ 等传输图像。对于分辨率较高、文件较大的图像可以使用有关模块压缩图像。报卡图像存入计算机的指定目录,通过程序给定一个惟一的记录号导入数据库。“图像导入”后,工作人员可以根据计算机显示的图像信息

输机。在数据管理时,每当显示一张卡片时,可以同时浏览原卡片的图像和最后修改卡的转换图像,便于发现输机错误,同时摒弃了庞大报卡柜和繁杂的找原始卡的过程。

病例管理视窗界面左上为当前卡的详细信息,右上为图像,下为表格显示的数据库。当卡片修改后,右中会显示修改卡的信息。当点击“关联卡”按钮时,与当前卡相关的卡片信息会显示在窗口下部,同时在右上可以浏览相关卡片图像,见图 3。



图 1 软件主视窗界面



图 2 登录窗口视窗界面

中国肿瘤登记数据管理系统 - [病例数据]

文件 编码 数据管理 文件管理 系统管理 窗口 窗口编排 病例输入 人口输入 帮助 当前用户:王庆生 =超级管理

《病例数据管理》 1 / 1168 直接输入新卡 删除当前报告卡 显示全部报告卡 显示关联报告卡

序号: 2 关联卡: 1 登记号: 20090004 报告状态: 0:未确定的 诊断日期: 姓名: 拼音: 身份证: 性别: 男性 出生日期: 年龄: 民族: 职业: 联系电话: -1 地区编码: 单位: -1 住址: ICD-03: 部位码: C34.9 形态码: 8000/3 / (分化码): ICD-10部位编码: C34.9 部位查询: C34.9 C34.9: 支气管或肺, 未特指《注释》 形态查询: 8000/3 8000/3: 肿瘤, 恶性《注释》肿瘤, 恶性, NOS/恶性肿瘤/癌症/未分类肿瘤, 恶性/恶性肿瘤, NOS 诊断依据: 2: X线、CT、超声波、内镜镜等 临床分期: T: N: M: 报告医院: 报告医生: 不详 报告日期: 多原发癌: 3: 三次原发 总数: 0 序数: 1 随访结果: 1: 存活 随访日期: 死亡地点: 5: 家中 死因: 1: 肿瘤 死亡日期: 婚姻状态=2; ICCC=12b; 检查状态=1; 个案重量=1; 生存月数=7; 备注:

开始图片 检索页面 现卡 1 原卡 1 现卡 2 原卡 2 现卡 3 原卡 3 现卡 4 原卡 4

病例报告模糊查询系统 清除选择 查询

关联卡号: 登记号: 住院号: 报告状态: 姓名: 拼音: 身份证号: 电话: 民族: 单位名称: 家庭住址: ICD-03: 部位: 形态学码: 分化码: ICD-10: 部位: 诊断依据: 卡来源: 临床分期: T分期: N分期: M分期: 报告医院: 报告医生: 多原发癌: 总数: 序数: 随访结果: 死亡地: 死因: 输机人员: 修改人员:

输机: 王庆生 2011/6/30 修改: 1赵先生 2011/7/15 查重: 8888/8/8 重编: 8888/8/8

序号	关联卡号	登记号	住院号	卡片状态	姓名	姓名拼音	性别	年龄	出生日期	身份证号	联系电话	民族
1	1	20090002	1212	2	邢		1	71	1937/5/10		-1	1
2	1	20090004	-1	2	孙		1	80	1928/11/8		-1	1
3	3	20090005	0115	2	王		1	70	1938/3/20		-1	1
4	4	20090006	47518	1	王		1	56	1953/11/26		-1	1
5	5	20090007	-1	1	杨		1	59	1949/8/7		-1	1
6	6	20090008	-1	1	赵		1	82	1927/2/26		-1	1
7	7	20090009	60503	1	孟		2	80	1928/12/27		-1	1
8	8	20090010	-1	1	李		2	62	1947/3/17		-1	1
9	9	20090011	-1	1	王		2	49	1959/5/29		-1	1
10	10	20090012	188636	1	李		1	71	1938/5/5		-1	1
11	11	20090013	2009412	1	邱		1	70	1938/8/3		-1	1
12	12	20090014	59275	1	刘		1	56	1952/12/12		-1	1
13	13	20090015	48066	1	邢		2	67	1941/11/19		-1	1

图3 病例数据管理视窗界面

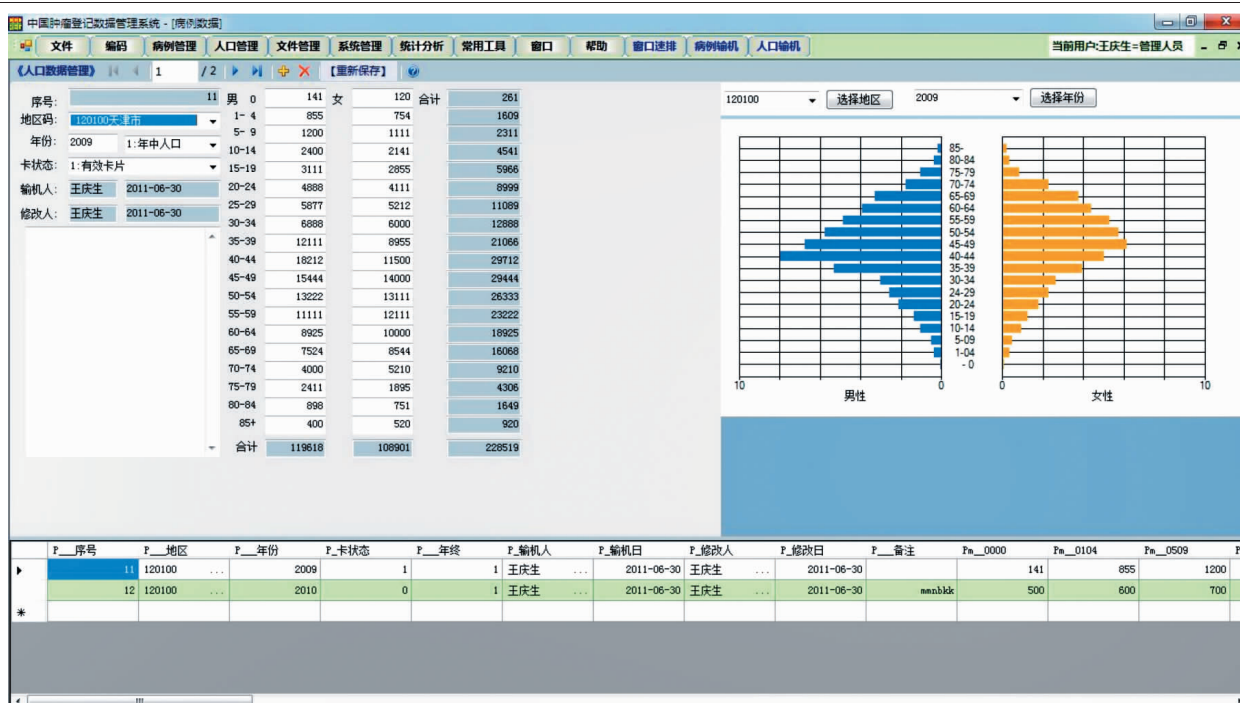


图4 人口数据管理视窗界面

人口管理视窗界面的左上为某一年的人口数据, 输机时合计项自动相加生成。在视窗界面的右上为人口金字塔形图。右中和窗口下部与病例视窗界面雷同。可以打开额外窗口显示各年龄组不同年份或不同地区的直方图, 见图 4。

病例和人口导出同时在同一个视窗界面进行, 可以选择限定输出范围。文本导入管理分别在两个视窗界面进行。

7 编码表与 ICD 编码模糊查询

编码表是数据管理标准化的重要组成部分。本系统可以通过两种方式打开编码表, “数据导出”子菜单和各个编码表菜单。前者不能修改, 但可以导出为 PDF、Excel 和 Word 格式。后者逐一打开各个编码表, 超级管理员权限可以修改有关编码。修改编码时一定要注意连续性和标准化, 不能随意改变。

在软件中还有一个重要功能, 就是 ICD-O-3 形态学编码、ICD-O-3 解剖学编码和 ICD-10 解剖学编

码的“ICD 模糊查询”模块。在需要模糊查询编码时, 可以输入 1~6 个关键词, 关键词可以是一个字或一个词组, 也可以是数字或者字母, 用空格分隔, 排列顺序不限。随着输入内容的变化, 附合条件的查询结果不断变化地缩小查询范围, 直到发现合适结果。此功能非常有助于对 ICD 编码不熟悉的工作人员快速找到合理编码, 也有助于资深工作人员正确判断不常见肿瘤部位或少见形态学描述的真实定义。

8 统计与报表

在“统计分析”模块部分, 包括一般的频数表、交叉表格、各种发病率和死亡率, 见图 5。对于部分表格, 如发病率和死亡率提供不同格式的报表以满足不同需求。对于肿瘤登记工作人员, 本软件还包括质量控制有关的报表。本部分的模块将根据需求不断地增加和完善。

在此模块, 还包括数据“检查错误”的功能。数据查错包括数据的逻辑错误检查, 包括两个程序, 一个



图 5 统计与报表视窗界面

是调用 IACRtools 的有关程序,输出是英文的结果,另一个是中文输出,依据的规则主要源于 IACRtools。数据查错还包括另一模块,用于检查数据输入的不规范的错误等。查错程序运行后,病例记录中将变更查错日期为程序运行日期,用于下次选择时限定依据。

“病例数据查重”是另一重要功能模块。查重时可以选择不同查重方案。与查错程序一样,查重后,相应记录将变更查重时间。

统计模块运行时要先导出有关记录,统计程序运行时调用导出记录的文本文件,但不需要人为地选择文件。程序运行后,输出文件将在子窗口中打开,如果要保存有关文件,请使用“文件”菜单中的“另存文件”子菜单。

9 其他功能与特点

除了上述功能外,本系统还有如下功能:

①多窗口显示功能,可以同时显示多个子窗口,并方便地相互切换,还可以使用“窗口速排”快捷键同时显示所有打开的子窗口。

②文本输出快捷和方便,并在必要时“用 word 编辑”打开相应文档进行排版建立高质量、可编辑的输出文件。

③建立有关信息的保存目录“文件管理”,并通过系统菜单快速打开。

④建立“常用工具”的快捷打开方式,如打开“计算器”。

10 未来肿瘤登记信息智能化管理展望

中国肿瘤登记历经数十年的发展,也已取得长

足的进步。由于人力物力投入较大,伴随医疗体制改革和科技进步,肿瘤登记的理念和方法也要与时俱进。在这个进程中,多网融化的国家级医疗一体化计算机网络和覆盖城乡居民的基本医疗卫生制度是肿瘤登记事业发展中必须关注的问题。

作为我国居民的主要死亡原因和常见病,癌症的防治应整合到国家整体疾病控制体系之中。从中长期发展规划来看,国家级医疗一体化计算机网络既是癌症防治的重要一环,也是国家疾病控制的终极选择。

国家级医疗一体化计算机网络建设至少应该包括如下内容:①全国各个医疗单位统一的,或可以相互转换的数据库,实现理论上的资源共享,逐步扩大应用面,最大限度地实现信息共享与利用。②建立地区性和国家医疗相关信息中心,智能化地合并地区性和国家级与医疗有关的数据库,包括人口资料(出生与死因统计信息—公安系统),医疗纪录(疾病信息—卫生系统)和医疗与社会保障纪录(经济相关信息—社会保障系统),通过计算机动态地智能收集数据。③根据需求设立权限,各个专业的人员分别处理分析相关数据。

从目前来看,马上建立这个国家系统还不成熟,但对于一些发达地区,应该探索可行性和开展试点工作。

肿瘤登记软件逐步地向网络版智能化数据采集系统的转化。中国肿瘤登记系统将成为高效的肿瘤登记系统。

致谢:本系统的开发过程中,河南省肿瘤医院孙喜斌教授、北京市肿瘤防治办公室王宁副主任、武汉市疾病控制预防中心段纪俊教授、浙江省肿瘤医院《中国肿瘤》编辑部夏庆民主任、河北医科大学肿瘤研究所贺宇彤教授、启东市肝癌研究所陈建国教授等专家提出了许多宝贵意见,在此表达衷心的感谢。