

主题数据库建设数据资源调查 和资源框架设计方法

医药卫生科学数据共享工程项目

资源体系建设课题组

二〇〇八年十一月

目 录

| | |
|-----------------------------------------------------|----|
| 概 述 | 3 |
| 一、基本概念 | 3 |
| 二、目的和目标 | 3 |
| 三、主题数据库建设任务和分工 | 4 |
| (一) 主题数据库建设总体任务 | 4 |
| (二) 任务分工 | 4 |
| 主题数据库建设科学数据资源调查和分析方法 | 6 |
| 一、适用范围 | 6 |
| 二、科学数据资源信息调查的原理和基本思路 | 7 |
| (一) 基本原理 | 7 |
| (二) 资源调查的基本思路 | 8 |
| (三) 建议优先考虑共享的科学数据资源 (资源遴选标准) | 8 |
| 三、资源调查和分析基本步骤 | 8 |
| (一) 选择资源信息库 | 8 |
| (二) 查询、摘要下载和建库 | 9 |
| (三) 文献计量学分析 | 9 |
| (四) 建立资源信息管理工具/系统 | 10 |
| (五) 资源确认调查和资源手册编写 | 10 |
| 四、资源调查和分析流程图 | 10 |
| 主题数据库资源框架设计方法 | 11 |
| 一、主题数据库资源框架目的和用途 | 11 |
| 二、主题数据库框架设计原则和分类依据 | 11 |
| (一) 主题数据库资源框架设计原则 | 11 |
| (二) 分类对象和分类依据 | 11 |
| 三、主题数据库资源框架设计步骤 | 12 |
| (一) 项目确立 | 12 |
| (二) 调研方案拟定、培训和实施 | 12 |
| (三) 主题数据库资源框架构建 | 13 |
| (四) 主题数据库资源框架优化 | 13 |
| (五) 主题数据库资源信息分类 | 14 |
| (六) 主题数据库资源框架提交 | 14 |
| 主要参考文献 | 15 |
| 附录 1 推荐查询的文献数据库介绍 | 16 |
| 附件 2 通过 CBMdisc 获取各类基金资助的肿瘤科学研究信息方法 | 20 |
| 附录 3 国家科技成果网查询结果和分析 | 27 |
| 附录 4 通过 Web of Science 查询在 SCI 期刊上发表肿瘤研究的科技论文 | 31 |
| 附录 5 万方科技成果数据库查询和分析结果 | 36 |
| 附录 6 肿瘤主题数据库框架设计结果 | 37 |
| 附录 7 资源信息管理工具/系统 | 41 |

主题数据库建设科学数据资源调查和资源框架设计方法

概 述

一、基本概念

1、科学数据资源

科学数据资源是指科技活动或通过其它方式所获取到的反映客观世界的本质、特征、变化规律等的原始基本数据，以及根据不同科技活动需要，进行系统加工整理的各类数据集，用于支撑科研活动的科学数据的集合。

2、主题

主题为围绕着某一项特定任务或活动进行数据规划和设计时，对其内容进行的系统归纳和描述。通常数据集主题应具有划分性和层级性，划分性是指主题间可通过不同的命名，将相同属性的主题归并在一起形成相同的类，将不同属性的主题区分开形成不同的类；层级性是指主题可被划分成若干子主题或子子主题。

3、主题数据库资源对象

具有特定主题的最小标识单元数据集。

4、主题数据库资源框架

为产生主题数据库资源分类。

5、主题数据库

通过特定资源框架组织的数据集集合。在一定的范围内，各数据集间具有语义(概念语义层)或语法（物理数据模式）一致性。

二、目的和目标

1、提供可行的和系统的科学数据资源调查方法，帮助科学数据资源组织者了解和掌握医药卫生科学数据资源存在的现状，建立持续稳定的科学数据资源收集渠道，解决科学数据共享工作中资源来源不足的问题，并确定资源遴选的基本准则，实现打造精品主题数据库的目标。

2、确定主题数据库资源分类框架建立方法，构建主题数据库资源分类框架，便于资源的组织和管理。

三、主题数据库建设任务和分工

(一) 主题数据库建设总体任务

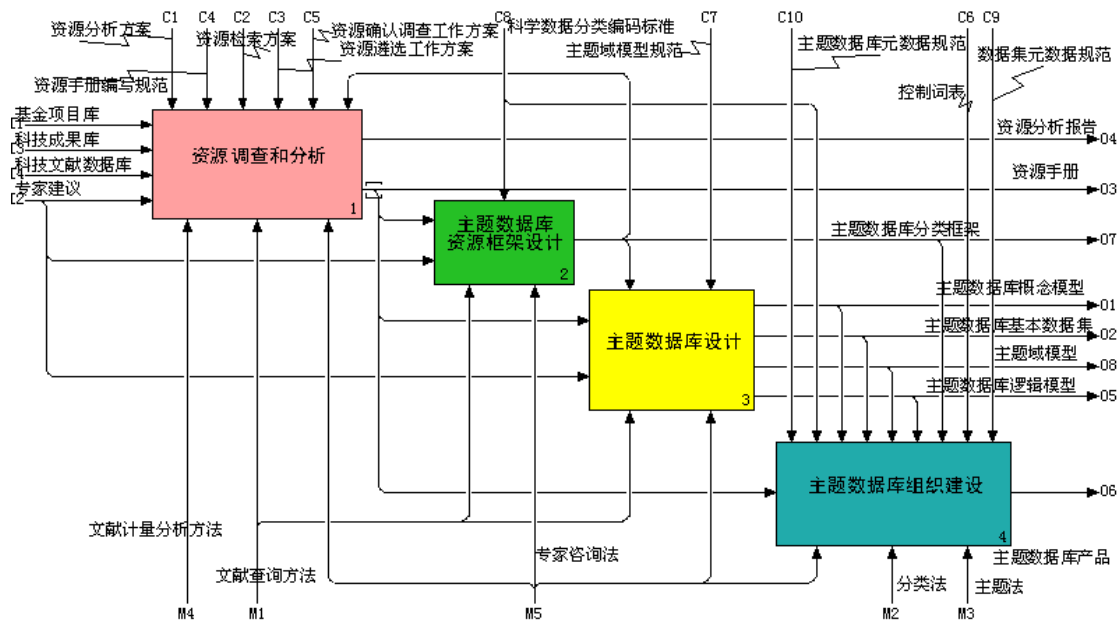


图 1、主题数据库建设总体任务

(二) 任务分工

1、主题数据库建设相关者

主题数据库规划设计涉及到总体组和各数据中心的资源、标准和平台组工作。

2、任务分工

资源组：资源调查和组织框架设计（语义关联性）

标准组：信息标准化（数据元和术语标准）

平台组：获取、加工、存储（数据模式一致性）

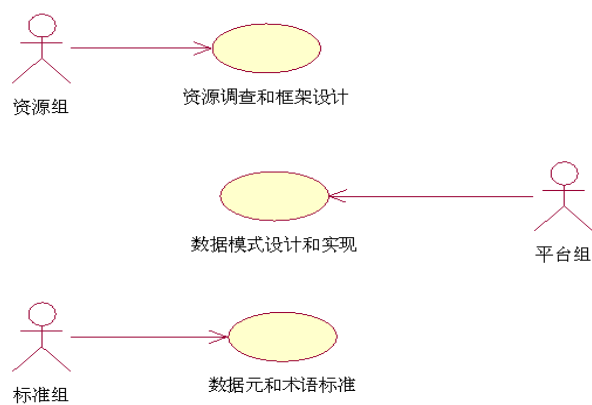


图 2 主题数据库建设总体任务分工

3、主题数据库资源建设任务分解

(1) 资源调查和分析

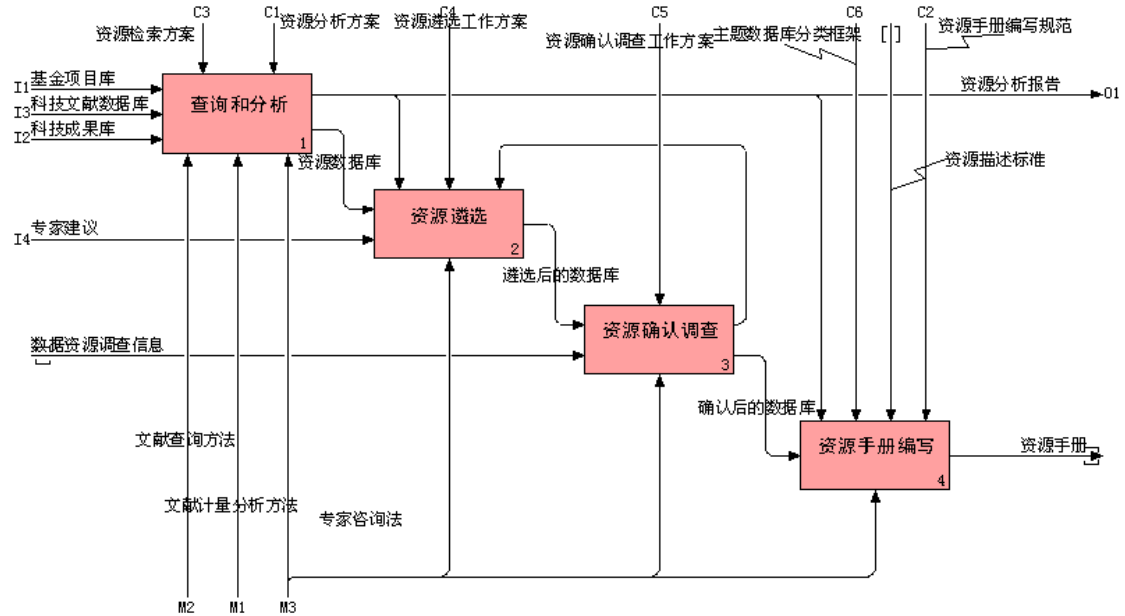


图 3、主题数据库资源规划和建设任务

(2) 资源框架设计

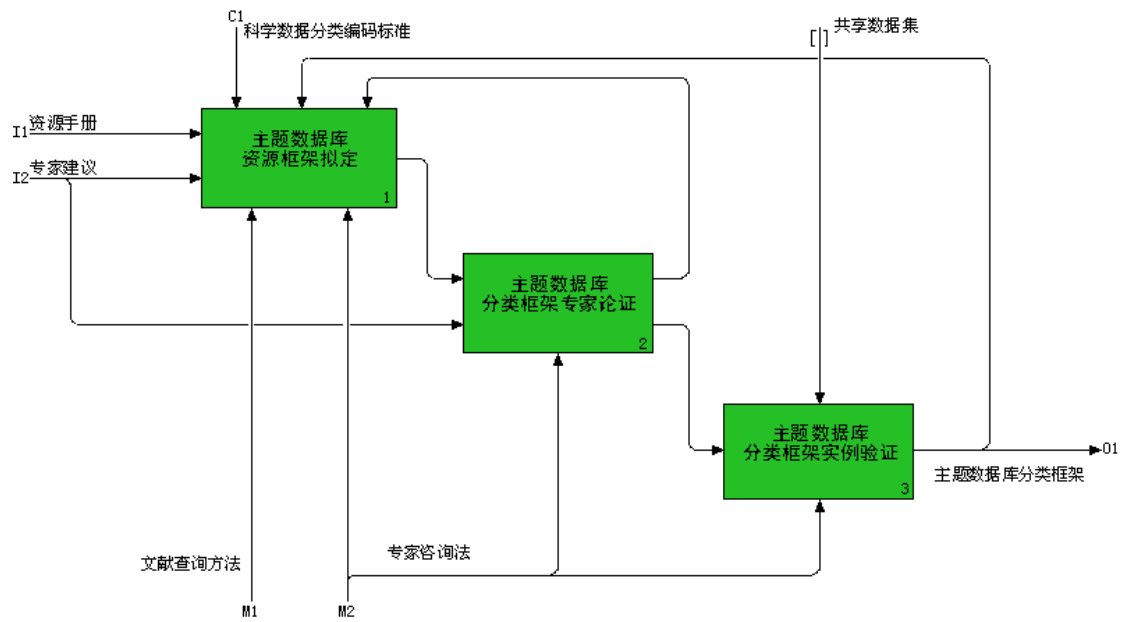


图 4、主题数据库资源框架设计任务

主题数据库建设科学数据资源调查和分析方法

一、适用范围

根据科技部对科学数据共享术语定义，科学数据是指人类在认识世界、改造世界的科技活动中所产生的原始性、基础性数据，以及按照不同需求系统加工的数据产品和相关信息。它既包括了社会公益性事业部门所开展的大规模观测、探测、调查、实验和综合分析所获得的长期积累与整编的海量数据，也包括国家科技计划项目实施与科技工作者长年累月科学实践所产生的大量数据。

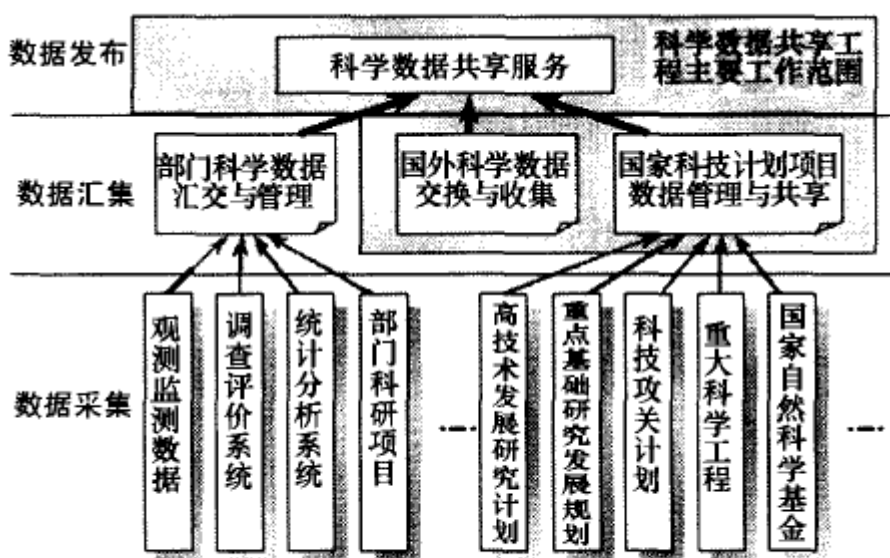


图5 《科学数据共享工程》的主要工作范围

从当前我国科学数据的产生和分布情况看，海量科学数据的采集、管理和维护，主要是通过政府投资完成的，是分布在单位乃至个人手中的国有资产。其所有权是国家所有，非数据生产单位或个人所有。我国当前情况下科学数据的获取和积累主要是靠政府投资完成的，它包括：

- 政府部门专项计划所产生的，维持其业务系统运行的专业数据；
- 国家各类科技计划、专项科技活动所产出的专业数据；
- 科研院所、高等院校所积累的专业数据；
- 国际科技数据组织、中国分支机构所拥有的国际科学数据资源；
- 科学家个人手中的科学数据等。

我国现阶段科学数据共享工程主要工作围绕三方面进行工作，第一以政府

部门的资源开放与共用为突破口；第二、解决国家重大科技计划项目数据的管理与共享；第三充分利用国际科学数据信息资源。¹

本科学数据资源信息调查方法主要针对国家和地方科研计划项目所产生的科学数据。对于政府部门和业务单位的科学数据资源信息获取，建议采用信息资源规划方法，按照拟定的信息资源汇交管理办法执行；可暂参照相关部门信息资源调查方法，如《北京市政府部门信息资源目录编制指南》、《中国疾控中心信息资源目录编制工作实施计划》²等方案中规定的资源调查方法。

二、科学数据资源信息调查的原理和基本思路

（一）基本原理

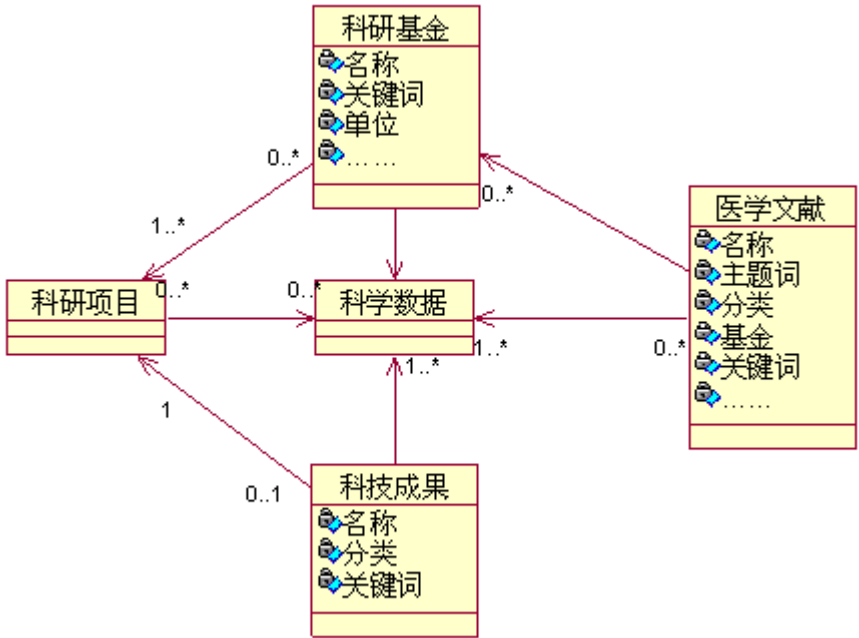


图 6、科研基金、科研项目、文献、科技成果与科学数据之间的关系

1、科研基金通过支持科研项目产生科学数据；因此通过科研基金信息可能追踪到科学数据；

2、通过分析科学数据，产生报告和论文，在学术期刊发表，其摘要被收录到题录文献库、全文库、引文库。因此通过查询相关文献库，可能帮助追踪到相关的科学数据；

3、一部分科研结果可能被评定为科技成果，并被收录到相关科技成果库中，

¹ 黄鼎成，李晓波，王卷乐. 浅谈科学数据共享工程建设的战略取向. 中国基础科学, 2005.7(5), 29-35.

² 文件下载: <http://www.phsciencedata.cn/ActDream/UserFiles/File/20071108info.doc>

通过查询科技成果库有可能追踪到相关的科学数据；

科学数据资源信息调查最终目的是获取科学数据（集）。通俗地说，就是能够告诉资源组织者在哪儿（where），在谁（who）的手中，有什么样（what）的数据集（data set）。

（二）资源调查的基本思路

1、首先收集有关科研基金、文献题录、科技成果等信息，建立科学数据资源追踪信息库。其中内容信息如摘要、正文、主题词、分类等可帮助确定主题；其归属信息可帮助确定可能存在的科学数据资源的生产/拥有者和单位。

2、根据科学数据资源追踪信息，对生产/拥有者和单位可能拥有的科学数据资源进行直接的追踪调查（如问卷、电话等），收集科学数据资源信息，对科学数据资源进行初步描述，形成科学数据资源手册。

3、数据中心参照科学数据资源手册进行有的放矢的数据采集、加工和服务，最终形成可共享的科学数据。

（三）建议优先考虑共享的科学数据资源（资源遴选标准）

- 1、各类基金资助的课题所产生的科学数据；
- 2、支持获得科技成果的科学数据；
- 3、上述两类科学数据中属于研究热点和影响程度高（如支持高分的 SCI 论文发表）的科学数据；
- 4、本领域内科技成果或论文多的权威机构或科学家个人拥有的科学数据；

三、资源调查和分析基本步骤

（一）选择资源信息库

1、科技成果数据库

通过查询各类科技成果数据库，可获得各类科技成果的内容、持有单位和个人，便于进一步进行科学数据资源的追踪。

（1）万方科技成果数据库

包括中国科技成果库、中国重大科技成果库、科技成果交易库、科技成果

精品数据库、国家级科技授奖项目数据库、科技决策支持库等 7 个成果库。

(2) 清华同方科技成果库

(3) 国家科技成果网

2、基金项目数据库

通过查询各类基金单位的数据库，了解各类基金历年来所资助的项目名称、研究内容、研究单位和个人。主要类别包括：

- 科技部科技项目：包括科技部科技基础性工作专项、社会公益研究专项、科技攻关、863 项目、973 项目等；
- 国家自然科学基金项目：国家自然科学基金资助项目等；
- 部委或部门项目：中科院项目、教育部项目、卫生部项目、其它部委项目等；
- 省市和地方项目
- 国际合作项目
- 自筹资金项目

3、科技文献库

国内：中国生物医学文献数据库(CBMDisc)、CMCC、清华同方 CNKI、维普、万方数据库等。

国外：Pubmed、Web of Science、Medline(silverplate 光盘版, Ovid, Ebsco 等)；

推荐使用中国生物医学文献数据库(CBMDisc)，通过所提供的基金支持情况，不仅可以获得各类基金所支持的科学研究产生的科技文献，而且还能掌握各类科研基金产生的科学数据所在单位和个人信息，便于进一步追踪。

(二) 查询、摘要下载和建库

根据不同文献库的特定，根据不同的主题要求，建立相应的检索策略，将查询获取的信息下载，并转换成结构化的数据库。

(三) 文献计量学分析

采用文献计量学方法，利用统计软件描述性分析：总体情况、时间趋势、领域分布（基础研究、临床研究、预防控制和中医药等）、学科分布、地域和机构分布、内容分布（分类、主题词）等。

分析性：共词分析、聚类分析

（四）建立资源信息管理工具/系统

利用信息技术建立单机或网络化的科学数据资源信息管理工具/系统，提供数据的录入、批量导入、导出、分类、查询、更新修改等功能，便于资源信息管理和利用。

（五）资源确认调查和资源手册编写

对象：资源拥有单位、资源拥有者；

调查内容：了解资源单位、资源拥有者的所拥有的数据集基本情况；

资源确认调查方式：问卷调查、电话咨询、网络调查等；

资源手册编写：根据确认调查获得的科学数据资源的数据集信息，撰写资源手册。

四、资源调查和分析流程图

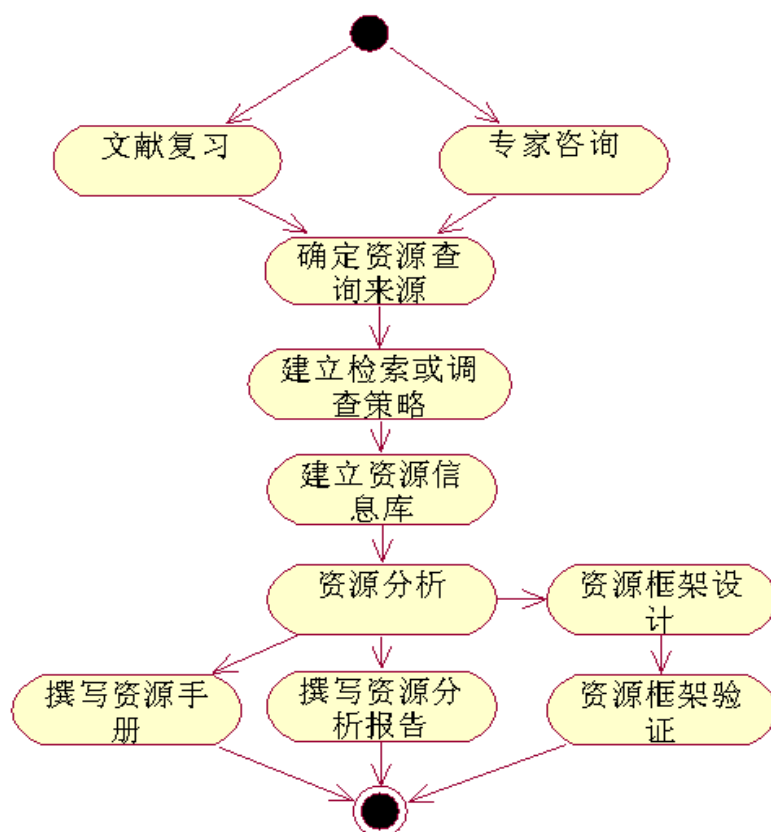


图 7 资源调查基本步骤

主题数据库资源框架设计方法

一、主题数据库资源框架目的和用途

主题数据库资源框架设计的目的是更好为科学数据资源组织和收集服务。主要用途包括：

- 1、作为资源规划目录，帮助科学数据资源的组织者，能够围绕特定的主题，系统性地采集科学数据。
- 2、提供信息导航服务，能够形成目录或树状的索引，方便使用者快速查找到相应的资源。
- 3、提供信息查询服务，分类条目能够作为词条，作为信息顺序编排和查询指针。
- 4、提供信息组织工具，把不同来源的科学数据信息通过逻辑方式进行整合。

二、主题数据库框架设计原则和分类依据

（一）主题数据库资源框架设计原则

按照《卫生信息数据集分类与编码的基本规则》进行。

（二）分类对象和分类依据

1、分类对象

主题数据库资源框架中的分类基于数据集。主题数据库资源框架设计任务需要将涉及到的一个或多个对象的数据集，进行抽象和归纳，以获得相同的主题。

2、分类依据

构建主题数据库资源框架建议从分析相关的业务活动入手，将业务活动的目标或功能作为分类依据。业务活动分析方法可直接借鉴软件工程学中业务分析的方法，如科学数据资源规划(IRP)业务梳理的方法等。

基于业务活动对数据集进行分类，具有很强的操作性。这是由于：

- （1）特定数据的产生，总是与具体的科研活动（业务活动）密切相关；
- （2）业务活动具有划分性和层级性，即大的业务活动，可再分成一系列粒度更小的活动；反之，小粒度也可再重新组合成新的或大粒度的活动；

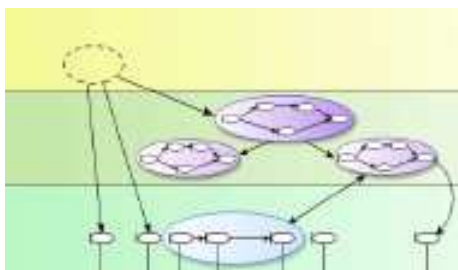


图 8 科学活动可分性和重组织性

三、主题数据库资源框架设计步骤

（一）项目确立

主题数据库的建设由医药卫生科学数据共享组总体组负责，由资源组统一规划，通过专家可行性论证后，纳入到科学数据共享的年度计划，在报批科技部并得到批准后才能立项。

（二）调研方案拟定、培训和实施

调查方案内容包括选择调研对象、调研内容，调研方法和途径、调研日程表、复查时间表等。需要对框架设计组和参与框架设计的管理人员进行适当的培训和指导。

调查内容：

- 1) 本主题领域内有关科学活动规划资料，包括科学活动的任务、目标、战略、计划等。
- 2) 本主题科学活动的组织结构的资料，即承担或组织整个科学活动的科研或业务实体及其分支机构等。
- 3) 本主题有关科学活动的资料，最好是科学活动结构图，如果没有，参照 IT 业务活动分析方法，绘制科学活动结构图，通过科学活动结构图把相应科学数据对应起来，揭示中间重复的科学活动，重组的科学活动。
- 4) 本主题领域内现有的数据资源的信息。
- 5) 本主题现有系统环境资料，即支持本主题各项科学活动的计算机信息管理系统及其采集和存储的数据。

调查方式或信息获取主要途径：

- 1) 通过审查文档资料，如科研规划、各种报表等；
- 2) 通过对代表性的科学数据产生的机构、生产者或拥有者等进行调研；
- 3) 通过对科技成果、科研基金、文献信息、科研项目等科技信息查询。根据

查询结果，建立相应的资源信息数据库。

（三）主题数据库资源框架构建

主题数据库资源框架构建采用科学数据资源规划(IRP)业务梳理的方法。采用通过对业务梳理分析,将主题数据库资源框架与业务模型(Business Model)统一起来,系统地、本质地、概括地把握科学活动的业务功能结构。

采用“职能域—业务过程—业务活动”这样的 3 层结构来梳理业务。具体步骤:

- 1) 把组织结构资料作为信息源,利用有关的软件工具建立组织层次图;
- 2) 建立科学活动业务基本模型,确定高层次的业务功能;
- 3) 将高层业务功能图进一步细化为底层业务功能图;
- 4) 业务功能映射到组织单元以后,即形成过程/组织 C-U 矩阵;
- 5) 把过程的输入、输出信息进行综合和分解从而得到数据类(相当于主题数据库),建立数据类/过程 C-U 矩阵。

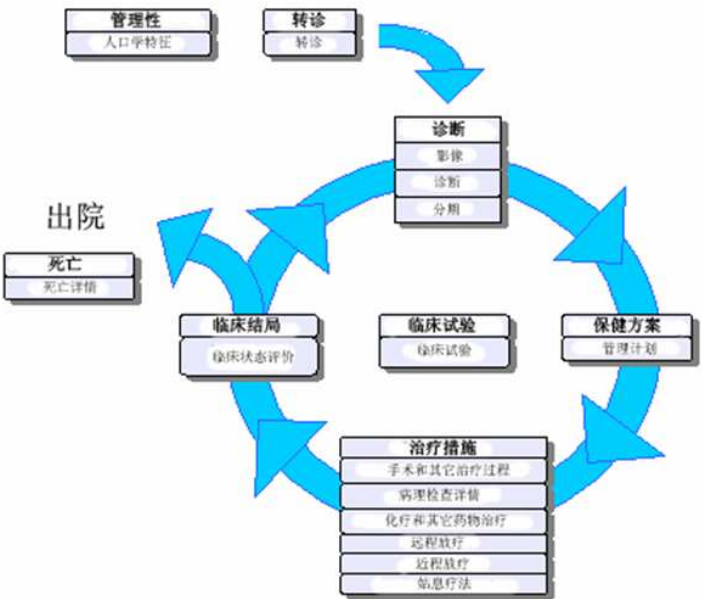


图 9 肿瘤主题数据库临床科学数据资源框架举例

（四）主题数据库资源框架优化

主要手段包括:

- 1) 依据现有的数据资源的信息进行调整;
- 2) 参照现有的信息管理系统及其数据进行调整;

3) 专家咨询。

(五) 主题数据库资源信息分类

利用主题数据库资源框架，对调研阶段形成的资源信息数据库进行分类。

(六) 主题数据库资源框架提交

主题数据库资源框架设计单位（资源组）向主题数据库组织和实现方（平台组）提交完整的主题数据库资源框架及其相关的数据资源信息，指导主题数据库建设中规划、任务分配、数据采集、分类和组织。主题数据库资源框架设计作为科学数据共享信息战略规划基础步骤，并根据对主题数据库建设工作的需求认识的深入，不断进行讨论和修正。

主要参考文献

1. 中国科技部. 科学数据共享工程技术标准化: 科学数据共享概念与术语—第 2 部分: 术语. 2005.
2. 孙九林. 科学数据资源与共享. 中国基础科学 2003; (1), 30-33.
3. 黄鼎成. 科学数据资源的管理. 中国基础科学 2006; 8(6), 20-24.
4. 黄鼎成, 李晓波, 王卷乐. 浅谈科学数据共享工程建设的战略取向. 中国基础科学 2005;7(5), 29-35.
5. 刘德培, 尹岭. 中国医药卫生科学数据共享实践. 计算机应用与软件 2006;23(10):1-3.
6. 高复先. 信息资源规划—信息化建设基础工程. 北京: 清华大学出版社; 2002.
7. 北京市政府部门信息资源目录编制指南(征求意见稿).2006
8. 中国疾病预防控制中心信息中心. 中国疾控中心信息资源目录编制工作实施计划.2007. <http://www.phsciencedata.cn/ActDream/UserFiles/File/20071108info.doc>

附录 1 推荐查询的文献数据库介绍

第一类：文摘数据库

国内

1、中国生物医学文献数据库(CBMDisc) (光盘版)

该数据库收录了中国科学院医学信息研究所编制的《中文科技资料目录 医药卫生》的 700 多种中国医学期刊以及汇编、会议论文的文献题录，94 年扩大收录范围并加入文摘，以光盘形式每月更新。

中国生物医学文献数据库(CBMDisc) 是中国医学科学院医学信息研究所开发研制的综合性医学文献数据库为国内最大也是最具权威的中文生物医学文摘索引型数据库。该库收录了 1978 以来 1600 多种中国生物医学期刊以及汇编、会议论文的文献题录，年增长量约 26 万条。1989 年以后的题录与维普全文数据库链接。涵盖了《中文科技资料目录(医药卫生)》、中文生物医学期刊目次数据库 (CMCC) 中收录的所有文献题录。

2、CMCC 文摘数据库

《中文生物医学期刊文献数据库》(Chinese Medical Current Contents, 简称 CMCC) 是解放军医学图书馆研发的医学期刊文献文摘型数据库。始创于 1994 年，收录了 1994 年以来国内正式出版发行的生物医学期刊约 1400 余种，文献量约 350 万篇；并以每年 35 万篇的速度递增。文献内容涉及基础医学、中国医学、临床医学、预防医学、药学、医学生物学、医院管理及医学情报等多个学科。文献记录中包括题名、作者 (全部)、关键词、摘要、第一作者的单位、邮政编码、文献出处、文献类型、参考文献数、译文出处等多项信息。提供单机版、网络版，支持 20 个或更多的并发用户授权。

国外

1、Medline 光盘版

该数据库由美国国立医学图书馆编制，囊括了全世界范围内 40 多个语种的 3000 余种生物医学期刊，其中包括 30 余种中国大陆生物医学期刊，每年收录约 40 万条文献记录，每条记录包括：文献标题、著者、著者地址、出处、语种、文摘、主题词等，可以打印和套录。

2、PubMed

网址：<http://www.ncbi.nlm.nih.gov/pubmed/>

PubMed 检索系统由隶属于美国国立图书馆 (NHL) 的国家生物信息中心 (NCBI) 开发。通过 PubMed 检索系统, 能免费检索到比 MEDLINE 收录范围更大的生物医学文献信息, 并且具有以下优点: 能上网即可获取该系统的相关信息, 检索功能强, 结果新, 可免费检索 MEDLINE、PREMEDLINE, 部分记录还可链接到全文。PubMed 收录比 MEDLINE 范围广, 在 MEDLINE 数据标引前的最新题录, 提供电子原文链接 (部分可免费获取全文)。

第二类: 引文数据库

ISI Web of science

网址: <http://www.isiknowledge.com>

科学引文索引是学术界公认权威的科技文献检索工具, 内容涵盖自然科学、工程技术、生物医学等 150 多个学科领域。目前该库有两种版本, 扩展版和核心版。扩展版收录期刊 5,900 多种。核心版 (印刷或光盘版) 收录 5900 种期刊中的 2100 种期刊。每周新增数据 19000, 新增参考文献 423000 条。

第三类: 科技成果数据库介绍

1、国家科技成果网

网址: <http://www.tech110.cn/tech/front/view/production/html/index.jsp>

目前包含 400,000 项科技成果, 通过全国省、部一级科技成果管理机构审核、认定、上报, 或由重点科研院所、211 重点工程院校直接报送, 并以每年 50,000 项的数量得到更新和追加。科技成果来源于 31 个省、自治区、直辖市, 16 个计划单列市及副省级城市, 以及 31 个国务院有关部门, 涵盖全国重点院所、重点大学、主要高新技术企业。最近 5 年的科技成果达到 200,000 项, 国家级奖励的科技成果共 2300 多项。全部科技成果涵盖国民经济各个行业, 其中以新技术、新产品为主要体现形式, 所占比重约 45% 和 20%。

2、万方科学成果系列数据库

网址: <http://www.wanfangdata.com.cn/>

包括中国科技成果库、中国重大科技成果库、科技成果交易库、科技成果精品数据库、国家级科技授奖项目数据库、科技决策支持库、中国医药科技成果数据库等成果库。

其中中国医药科技成果数据库由中国科技信息研究所提供, 收录 1986 年以来

生物、医药领域新技术、新产品、新工艺、新材料、新设计等技术成果项目。数据来自历年各省市部委鉴定后上报国家科技部的科技成果及星火科技成果。该库为国家指定的科研成果查新数据库。

第四类：中文期刊全文数据库

(1) 中国医院知识仓库

网址：<http://www.chkd.cnki.net/>

《中国医院知识仓库》总库（简称 CHKD 总库）是专门针对医务人员临床疑难病症诊断治疗，医学科研项目选题、设计、撰写论文、成果鉴定、医院管理人员决策经营，医院科技项目查新和科研绩效评价，医务人员继续医学教育等多方面的知识信息需要，开发的专业知识仓库，是 CNKI 系列数据库的重要专业知识仓库之一。是我国第一套具有主题词、分类号智能检索的医学专业中文全文数据库系统。由《CHKD 期刊全文数据库》、《CHKD 博士学位论文全文数据库》、《CHKD 优秀硕士学位论文全文数据库》、《CHKD 重要会议论文全文数据库》及《CHKD 重要报纸全文数据库》构成。2004 年被卫生部科教司评定为查新数据库。内容范围：基础医学、临床医学、预防医学、中国医学、药学、特种医学、生物科学、经营管理、图书情报、计算机及应用、医学教育与外语学习等多个学科专业。

(2) CHKD 期刊全文数据库

网址：<http://www.chkd.cnki.net/>

收录了我国公开出版发行的生物医学类专业期刊和相关专业期刊，整刊 1500 多种，部分刊 4000 多种。累计收录 1994 年至今的文献量达 430 多万篇，每年新增 50 多万篇。内含各科进展、动态、综述、概况类文章 10 万余篇，临床病例病案报道 86 万多篇，最新治疗措施 44 万余例。

(3) 《CHKD 博硕士学位论文全文数据库》

网址：<http://www.chkd.cnki.net/>

是目前国内相关资源最完备、收录质量最高、连续动态更新的中国生物医学类博硕士学位论文全文数据库，收录 1999 年至今全国博士培养单位的优秀博、硕士学位论文近 8 万篇。

(4) 万方期刊数据库

网址：<http://www.wanfangdata.com.cn/>

万方资源系统的文献服务是以拥有的数量巨大、种类繁多、专业覆盖范围广泛的书本式和电子版科技文献资源为基础,结合先进的网络系统、计算机数据库和检索技术而发展起来的快速、简单、有效的网上文献服务系统。通过 Internet 和 WWW,提供科技文献目录的网上查询和中外原文文献索取服务,并可以查看 1998 年以来 16 个专业上百种杂志的全文。既可以由杂志直接进入阅读,也可以按关键词、作者、刊名检索后再阅读,同时可打印、套录。

(5) 中文科技期刊数据库

网址: <http://www.cqvip.com/>

重庆维普资讯有限公司的主导产品《中文科技期刊数据库》是经国家新闻出版总署批准的大型连续电子出版物,收录中文期刊 12000 余种,全文 2300 余万篇,引文 3000 余万条,分三个版本(全文版、文摘版、引文版)和 8 个专辑(社会科学、自然科学、工程技术、农业科学、医药卫生、经济管理、教育科学、图书情报)定期出版。该库的时间跨度、收录期刊种类及文献量在国内同类产品中都首屈一指。

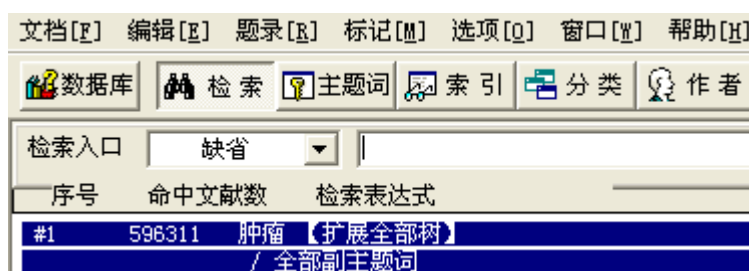
附件 2 通过 CBMdisc 获取各类基金资助的肿瘤科学研究信息方法

一、目的

获取 CBMDISC 中 CBMDISC94-CBMDISC08 数据库中有关原创性肿瘤研究论文记录，只包括原始性研究论文，不包括病例报告、综述和其他二次文献。

二、建立检索策略

1.使用“肿瘤”主题词查询



2.按照“R73”肿瘤学分类查询



3.限定去除病例报告、综述等文献类型

4.只选择所有的基金资助的文献摘要

三、摘要套录

根据需要，确定获取字段，将文献题录套录下载。

1、获取检索字段

| | | |
|---|------|------------------|
| 1 | 分类号 | 学科分布 |
| 2 | 标题 | |
| 3 | 文献类型 | |
| 4 | 著者 | 作者分布（是否可把作者单独提取） |
| 5 | 著者单位 | 机构分布 |
| 6 | 国省市名 | 地域分布 |
| 7 | 文摘 | 内容分析 |

| | | |
|----|------|-----------|
| 8 | 基金 | |
| 9 | 出处 | 期刊分布/时间分布 |
| 10 | ISSN | |
| 11 | 关键词 | |
| 12 | 主题词 | 内容分析 |

2、逐步将所有文摘进行下载（每次只能下载 500 条记录）

三、套录文档合并和转换

利用 SAS 数据过程将文本文档转换成结构化数据。

1、原始文本文档

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| CBMdisc97 | 1/990 |
| 流水号: 2002059580 | |
| 分类号: R342.3; R394; R394.2; R735.202.6; R735.350.22; R735.350.26; R814.3; R977 | |
| 标题: nm23 基因表达与胃癌、大肠癌 DNA 含量的关系 | |
| 英文标题: Study of the relationship between nm23 gene expression and nuclear DNA content in human gastric and colorectal carcinomas | |
| 文献类型: 英文摘要 | |
| 著者: 王川; 张祥福; 杨发端; 卢辉山; 殷凤峙 | |
| 著者单位: 福州福建医科大学附属协和医院 350001 | |
| 国省市名: 福建 | |
| 中文摘要: 探讨 nm23 基因产物/NDPK 表达与胃癌、大肠癌细胞核 DNA 含量的关系。方法: nm23 基因产物/NDPK 表达阴性、弱阳性及强阳性的胃癌、大肠癌共 57 例(胃癌各 11 例、大肠癌各 8 例),使用 DNA 图像分析技术检测癌细胞核 DNA 含量。结果: nm23/NDPK 表达阴性的胃癌或大肠癌细胞核 DNA 含量显著高于 nm23/NDPK 阳性表达者(分别为 $P<0.05$, $P<0.001$);在 DNA 倍体分布形式上, nm23/NDPK 阴性的胃癌和大肠癌中二倍体细胞的检出率较低(分别为 $P<0.05$, $P<0.001$);而非整倍体细胞检出率则显著高于 nm23/NDPK 阳性表达者(分别为 $P<0.05$, $P<0.001$)。结论: nm23 基因产物/NDPK 失表达与胃癌、大肠癌非整倍体细胞形成有关。 | |
| 著者文摘: Author | |
| 参文数: 9 | |
| 资助类别: 福建省科委重点项目基金资助 | |
| 资助编号: 95-Z-133 | |
| 出处: 中国体视学与图像分析 1997.03.15; 2(1): 36-38 | |
| ISSN : 1007-1482 | |
| 国内代码: 11-3739/R | |
| 内部代码: Z118 | |
| 出版地: 北京 | |
| 关键词: DNA 含量; 基因表达; 结直肠肿瘤; 图像分析技术; 胃肿瘤 | |
| 主题词: DNA/分析; *基因, 肿瘤抑制; *基因表达; 结直肠肿瘤/*遗传学; 结直肠肿瘤/ | |

| |
|-------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>病理学; 图像解释, 计算机辅助; 胃肿瘤/*遗传学; 转录因子</p> <p>特 征 词: 人类</p> <p>CSICI : 1007-1482(1997)2:1<36:JYBDYWADCA>2.0.TX;1-T</p> <p>更新日期: 19971231</p> |
|-------------------------------------------------------------------------------------------------------------------------------------------------|

2、SAS 转换程序代码

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <pre> %macro literature(pathway=); data aa; infile &pathway. DLM='09'x DSD LRECL=80 TRUNCOVER FIRSTOBS=1; length a \$80. ; input a ; a=compbl(a); run; data aa; set aa; drop a; length b \$10 idnum 8; retain b idnum; if count(a, "CBMdisc")>0 then idnum=_n_; if count(a, ": ")>0 then do; b=substr(a, 1, find(a, ": ")-1); c=substr(a, find(a, ":")+2,length(a)-find(a, ": ")-1); end; else c=trim(a); sn=_n_; run; proc sort data=aa; by idnum b sn; run; data aa; set aa; by idnum b; length d \$800.; retain d; if first.b and last.b then d=c; </pre> | <pre> when ('国内代码') id='Countycode'; when ('流水号') id='Serial'; when ('内部代码') id='Innercode'; when ('特征词') id='Character'; when ('文献类型') id='Type'; when ('英文标题') id='E-title'; when ('中文摘要') id='Abstract'; when ('主题词') id='MESH'; when ('著者') id='Author'; when ('著者单位') id='Address'; when ('著者文摘') id='A_SO'; when ('资助编号') id='Fundnum'; when ('资助类别') id='Fund'; when ('国省市名') id='Province'; otherwise id="Unknown"; end; run; data aa; set aa; if id ne "Unknown"; run; proc sort data=aa; by idnum sn; run; proc transpose data=aa out=total ; id id; idlabel b; var d; by idnum; run; proc datasets gennum=all; delete aa; run; </pre> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <pre> else do; if first.b then d=c; else d=trim(d) c; end; if last.b then output; drop c; run; data aa; set aa; length id \$10.; b=compress(b); select (b); when ('CSICI') id='CSICI'; when ('ISSN') id='ISSN'; when ('标题') id='Title'; when ('参考文献') id='Reference'; when ('参文数') id='Refnum'; when ('出版地') id='Publisher'; when ('出处') id='Source'; when ('分类号') id='Category'; when ('更新日期') id='Update'; when ('关键词') id='Keyword'; </pre> | <pre> DATA total; SET total; DROP IDNUM _NAME_ A_SO Update CSICI Reference Innercode Character; RUN; proc sort data=total nodupkey; by Serial; run; %mend; %literature(pathway='J:\chuli\肿瘤基金论文\all.txt'); libname xx "D:\文献调查结果\cbmdisk"; data xx.total_fund; set total; run; </pre> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

3、结构化数据（举例）



| | A | B | C | D | E | F | G | H | I | |
|----|---------|------------------------|-----------|--------|---------|---------|------|----------|----------|----------------------------|
| | idnum | title | idnum1 | author | address | keyword | mesh | category | fund | abstract |
| 1 | 9900990 | 人卵巢癌裸鼠移植瘤传代生长的研究 | 198902236 | 魏丽惠 | 军事医学科学院 | 卵巢肿瘤 | 卵巢肿瘤 | 肿瘤学 | R73-352 | 中国科学院基金 |
| 2 | 9900989 | 单层培养的垂体生长激素分泌瘤组织 | 198902875 | 邓洁英 | 军事医学科学院 | 垂体肿瘤 | 垂体肿瘤 | 内分泌学 | R347.511 | 国家自然科学基金资助项目 |
| 3 | 9900988 | 垂体泌乳素瘤下丘脑多巴胺调节作用 | 198903805 | 夏补 | 军事医学科学院 | 垂体肿瘤 | 垂体肿瘤 | 血液学 | R347.512 | 国家自然科学基金 |
| 4 | 9900987 | 五种上皮性肿瘤中细胞增殖细胞的增殖 | 198903815 | 赵玉松 | 军事医学科学院 | 肿瘤学 | 肿瘤学 | 病理学 | R329.24 | 国家自然科学基金 |
| 5 | 9900986 | β-微管蛋白对表皮角质化细胞增殖的影响 | 198903876 | 黄敬 | 军事医学科学院 | 肿瘤学 | 肿瘤学 | 病理学 | R329.24 | 国家自然科学基金 |
| 6 | 9900985 | 用单克隆抗体对淋巴瘤浸润皮肤 | 199000565 | 顾绍棠 | 军事医学科学院 | 淋巴瘤 | 淋巴瘤 | 免疫学 | R331.125 | 国家自然科学基金 |
| 7 | 9900984 | 人胰腺癌表皮生长因子及受体研究 | 199000910 | 陈元方 | 军事医学科学院 | 胰腺癌 | 胰腺癌 | 免疫学 | R730.26 | 国家自然科学基金 |
| 8 | 9900983 | 人卵巢癌裸鼠移植瘤体内试验性研究 | 199001235 | 魏丽惠 | 军事医学科学院 | 卵巢肿瘤 | 卵巢肿瘤 | 肿瘤学 | R730.269 | 国家自然科学基金 |
| 9 | 9900982 | 抗人胰腺癌组织单克隆抗体的研制 | 199002044 | 袁世珍 | 军事医学科学院 | 胰腺癌 | 胰腺癌 | 免疫学 | R329.25 | 国家自然科学基金 |
| 10 | 9900981 | 激光加热治疗小鼠前胃腺转移癌 | 199003176 | 高孟林 | 军事医学科学院 | 激光 | 激光 | 免疫学 | R312 | R45 国家自然科学基金 |
| 11 | 9900980 | 遗传性部位特异性结肠癌：两个家系 | 199003178 | 邵永学 | 军事医学科学院 | 结肠癌 | 结肠癌 | 遗传学 | R735.35 | 国家自然科学基金 |
| 12 | 9900979 | 用原位杂交技术对肝细胞性肝癌 | 199003181 | 陈梅龄 | 军事医学科学院 | 肝癌 | 肝癌 | 免疫学 | R342.33 | 中国国家自然科学基金、七五攻关、应用生物标记的HBI |
| 13 | 9900978 | 人胰腺癌裸鼠移植瘤系的建立及其 | 199003183 | 陈杰 | 军事医学科学院 | 胰腺癌 | 胰腺癌 | 肿瘤学 | R73-352 | 国家自然科学基金和中国科学院 |
| 14 | 9900977 | 原位肝肿瘤高灌注注大量5-氟尿嘧啶 | 199003526 | 王颖 | 军事医学科学院 | 肝癌 | 肝癌 | 免疫学 | R735.7 | 广东省科委基金资助 |
| 15 | 9900976 | 急性白血病综合药敏试验临床意义 | 199003825 | 王寿康 | 军事医学科学院 | 白血病 | 白血病 | 免疫学 | R733.71 | 卫生部科学基金资助 |
| 16 | 9900975 | 肺癌患者肺的细胞亚群和自然杀伤 | 199003831 | 张德祥 | 军事医学科学院 | 肺癌 | 肺癌 | 免疫学 | R331.125 | 国家自然科学基金 |
| 17 | 9900974 | 核素标记抗人胃癌单克隆抗体RWS | 199003835 | 吴裕沂 | 军事医学科学院 | 胃癌 | 胃癌 | 免疫学 | R735.203 | 国家自然科学基金 |
| 18 | 9900973 | 大肠癌旁移植物瘤组织抗原免疫组 | 199003841 | 王强 | 军事医学科学院 | 大肠癌 | 大肠癌 | 免疫学 | R322.45 | 其它基金 |
| 19 | 9900972 | 抗胃癌单克隆抗体M7和M4-1在胃 | 199100584 | 金晓 | 军事医学科学院 | 胃癌 | 胃癌 | 免疫学 | R392.11 | 国家自然科学基金 |
| 20 | 9900971 | 肠管切缘粘膜蛋白组织化学变化 | 199100925 | 王强 | 军事医学科学院 | 胃癌 | 胃癌 | 免疫学 | R341 | R73 中国解放军医药卫生科研基金 |
| 21 | 9900970 | 几种皮肤肿瘤中角质细胞表达CKM511 | 199101311 | 何春萍 | 军事医学科学院 | 肿瘤学 | 肿瘤学 | 免疫学 | R329.24 | 卫生部科研基金 |
| 22 | 9900969 | 恶性肿瘤患者外周血单个核细胞中 | 199101585 | 陈毓仙 | 军事医学科学院 | 肿瘤学 | 肿瘤学 | 免疫学 | R329.24 | 国家自然科学基金 |
| 23 | 9900968 | 激素、生长因子对体外胃粘膜上皮 | 199101605 | 张启行 | 军事医学科学院 | 胃癌 | 胃癌 | 免疫学 | R322.44 | 中国科学院基金 |
| 24 | 9900967 | 7,12-二甲基苯并Thy-1初发突变 | 199101621 | 管理 | 军事医学科学院 | 胃癌 | 胃癌 | 免疫学 | R329.24 | 国家自然科学基金 |
| 25 | 9900966 | 地鼠腺瘤及正常腺瘤对1.06μm | 199101624 | 吴美娟 | 军事医学科学院 | 胃癌 | 胃癌 | 免疫学 | R322.41 | 国家自然科学基金 |
| 26 | 9900965 | Nd-YAG激光治疗口腔颌面部鳞癌 | 199101624 | 赵芬 | 军事医学科学院 | 口腔癌 | 口腔癌 | 免疫学 | R732.205 | 国家自然科学基金 |
| 27 | 9900964 | 1,25(OH)2D3对HL-60细胞的抑制 | 199102075 | 严蔚 | 军事医学科学院 | 肿瘤学 | 肿瘤学 | 免疫学 | R73-351 | 其它基金 |
| 28 | 9900963 | 人肺癌组织中超氧化物歧化酶活性 | 199102325 | 江朝光 | 军事医学科学院 | 肺癌 | 肺癌 | 免疫学 | R345.49 | 国家自然科学基金资助的课题 |
| 29 | 9900962 | 多巴胺和溴隐亭对体外垂体生长 | 199103165 | 邓洁英 | 军事医学科学院 | 垂体肿瘤 | 垂体肿瘤 | 免疫学 | R347.511 | 国家自然科学基金 |
| 30 | 9900961 | 六株人肺癌细胞系侵袭特性的比较 | 199103404 | 夏丽红 | 军事医学科学院 | 肺癌 | 肺癌 | 免疫学 | R341.7 | R 国家自然科学基金 |
| 31 | 9900960 | 99mTc-PMT肝细胞癌阳性显像诊断 | 199103415 | 陈绍亮 | 军事医学科学院 | 肝癌 | 肝癌 | 免疫学 | R341.7 | R 国家自然科学基金 |
| 32 | 9900959 | 吡喃酮(PMA)对人细胞白血病细胞 | 199200605 | 姚晓斌 | 军事医学科学院 | 白血病 | 白血病 | 免疫学 | R730.231 | 其它基金 |
| 33 | 9900958 | 125I标记的抗人胃癌单克隆抗体 | 199200615 | 赵芬 | 军事医学科学院 | 胃癌 | 胃癌 | 免疫学 | R341.7 | 国家自然科学基金 |

四、分析结果（以国家自然科学基金为例）

1、地区分布

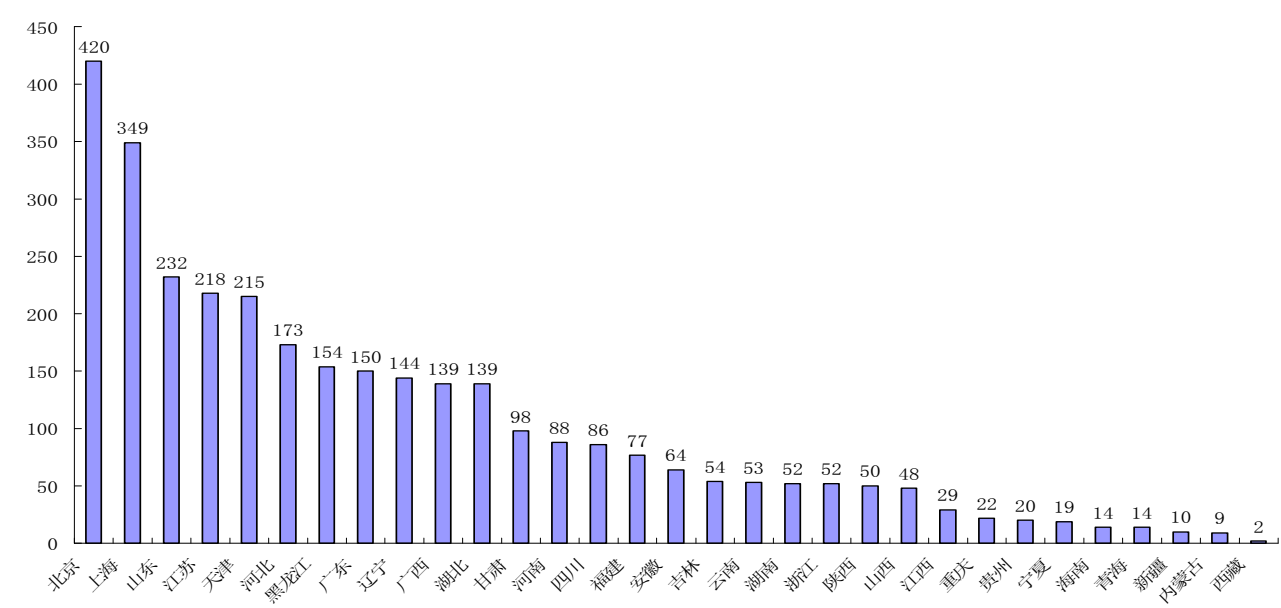


图 1 成果的地区分布

2、分类统计

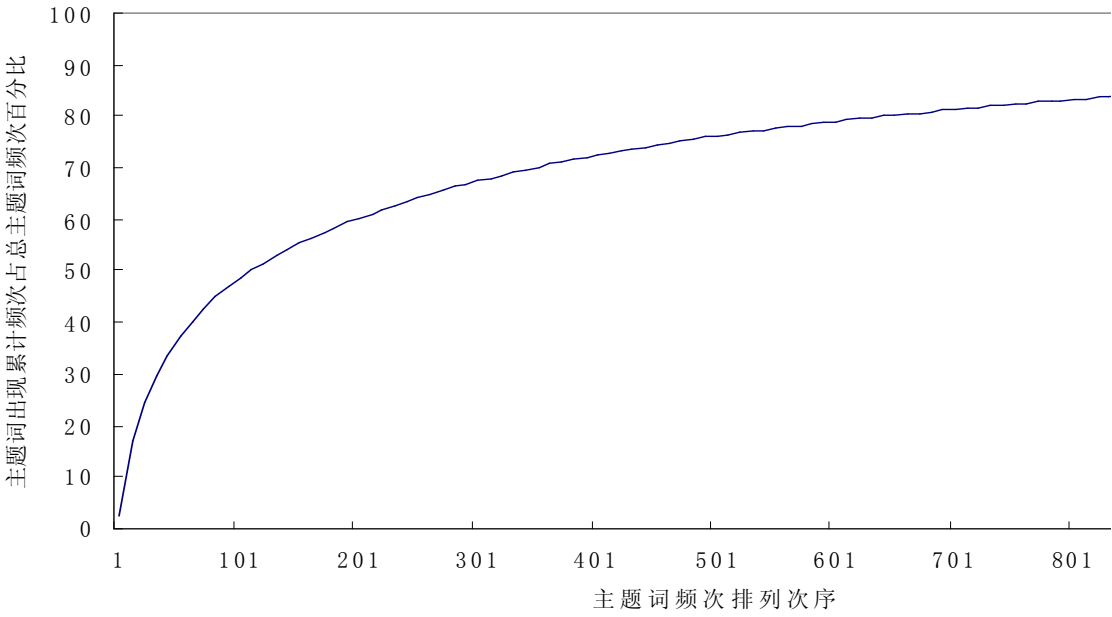
表 1 根据中图法分类统计结果（1）

| 中图法 | 类目一、二 | 类目三 | 类目四 | 类目五 | 文献数 |
|------------|-----------|-------------|-----------|------|------|
| R73 | 肿瘤学 | | | | 194 |
| R73-3 | 肿瘤学实验研究 | | | | 189 |
| R73-31 | | 肿瘤学调查与统计 | | | 7 |
| R73-35 | | 移植与组织培养 | | | 1 |
| R73-351 | | | 组织培养 | | 1445 |
| R73-352 | | | 移植与接种 | | 157 |
| R73-354 | | | 实验性肿瘤 | | 225 |
| R73-36 | | 治疗实验 | | | 5 |
| R73-362 | | | 生物疗法实验 | | 5 |
| R73-37 | | 肿瘤的转移与扩散 | | | 464 |
| | | | | | 1 |
| | | | | | 10 |
| R730.1 | 肿瘤的预防与控制 | | | | 23 |
| | | | | | 2 |
| R730.2 | 肿瘤病理学、病因学 | | | | 285 |
| R730.21 | | 肿瘤病理解剖学、组织学 | | | 212 |
| R730.23 | | 肿瘤病理生理学 | | | 37 |
| R730.231 | | | 肿瘤病因学、发生学 | | 90 |
| R730.231.1 | | | | 化学因素 | 388 |
| R730.231.2 | | | | 物理因素 | 25 |
| R730.231.3 | | | | 生物因素 | 90 |
| R730.231.9 | | | | 其他因素 | 443 |
| | | | | | 2 |
| R730.26 | | 肿瘤病理类型 | | | 279 |

表 2 根据中图法分类统计结果（2）

| 中图法 | 类目一、二 | 类目三 | 类目四 | 类目五 | 文献数 |
|----------|-------|-----------|--------|-----|-----|
| R730.261 | | | 上皮组织肿瘤 | | 936 |
| R730.262 | | | 间叶组织瘤 | | 90 |
| R730.263 | | | 造血组织瘤 | | 1 |
| R730.264 | | | 神经组织瘤 | | 449 |
| | | | | | 1 |
| R730.269 | | 肿瘤病理类型 | 其他组织瘤 | | 76 |
| R730.3 | | 肿瘤免疫学与血清学 | | | 159 |
| | | | | | 1 |
| R730.4 | 肿瘤诊断学 | | | | 39 |
| R730.43 | | 实验室诊断 | | | 32 |
| R730.44 | | 放射线、同位素诊断 | | | 11 |
| R730.45 | | 免疫诊断 | | | 1 |
| | | | | | 1 |
| R730.5 | 肿瘤治疗学 | | | | 203 |
| R730.51 | | 免疫疗法 | | | 9 |
| R730.52 | | 中草药疗法 | | | 5 |
| R730.53 | | 化学（药物）疗法 | | | 65 |
| | | | | | 5 |
| R730.54 | | 生物疗法 | | | 4 |
| R730.55 | | 放射疗法 | | | 33 |
| R730.56 | | 手术（外科）疗法 | | | 3 |
| R730.57 | | 激光疗法 | | | 2 |
| R730.59 | | 其他 | | | 9 |
| R730.6 | 肿瘤并发症 | | | | 2 |

3、主题词分析



所有 9313 篇包含主题词的论文中共提取主题词为 4274 个，累计出现频次为 54643，其中前 200 个主题词累计出现的频次为 32883，占总主题词累积频次的 60%，覆盖论文 97.2%

第一类：肿瘤疾病诊断类（44 个）

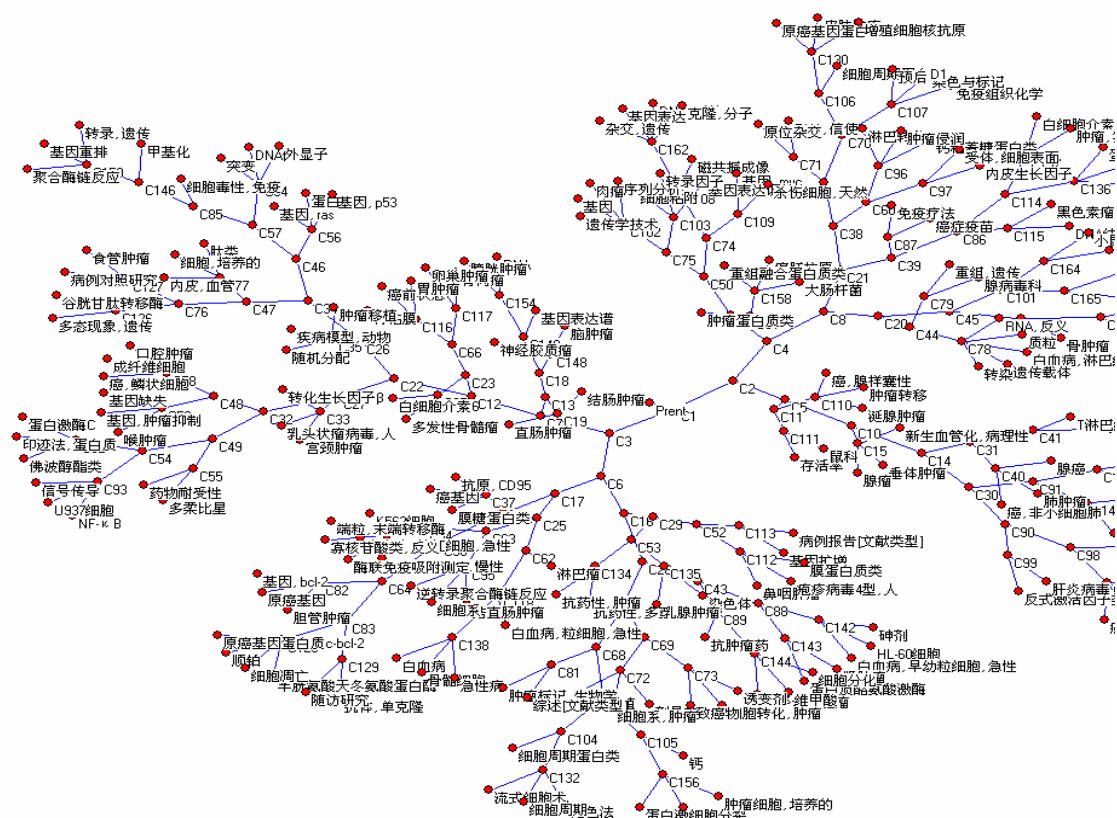
第二类：化学类（38）

 肿瘤遗传功能物质（18）

 肿瘤生长相关因子（5）

- 肿瘤免疫相关物质 (15)
- 第三类：解剖类 (18)
- 肿瘤组织病理 (4)
- 肿瘤细胞系 (11)
- 遗传物质载体 (3)
- 第四类：机制类 (21)
- 肿瘤细胞机制 (8)
- 肿瘤分子机制 (9)
- 肿瘤转移机制 (4)
- 第五类：环境因素类 (8)
- 第六类：研究方法和手段类
- 试验研究方法 (18)
- 动物模型 (4)
- 肿瘤人群研究方法 (5)
- 第七类：肿瘤诊断治疗类 (15)
- 肿瘤治疗 (13)
- 肿瘤诊断 (1)
- 肿瘤预后 (1)

4、共词分析



附录 3 国家科技成果网查询结果和分析

一、查询结果

1、总体情况

医药卫生分类共有 57765 条成果

2、数据结构形式

a) 简要形式（举例）

| 中药材真伪鉴别资料库 | | | | |
|------------|---------------|-------|---------|-------|
| 成果类型： | 年份：2005 | 成果水平： | 所处阶段： | 研究形式： |
| 关键词： | 中药材；真伪鉴别；资料库； | | 成果体现形式： | |
| 国科网分类： | 医药卫生；软件；I | | 行业分类： | |
| 学科分类： | 中药学；计算机软件 | | 高新技术分类： | |

成果摘要

数据更新

中药材真伪鉴别资料库收集了正品中药材原色图谱，同时也收集了该品种的常见混淆品的原色图谱，以及显微鉴别、快速鉴别、光谱与色谱的鉴别、检验标准等资料，非常适用于中药材检验工作人员的使用。设计的软件建立自动检索方式、所收录的中药材的原色图谱具有放大与缩小功能，还可实现全屏显示。收录的项目有形状鉴别、来源、显微鉴别、快速鉴别、光谱鉴别、常见混淆品鉴别、检验标准。共完成原色图片1029幅、178种快速鉴别方法、可鉴别467种常用的中药材。

b) 详细形式（举例）

成果概况：中药材真伪鉴别资料库

成果类型：

年份：2005

成果水平：

所处阶段：

研究形式：

成果编号：00000000000000439088

成果属性：

关键词：中药材；真伪鉴别；资料库；

成果体现形式：

国科网分类：医药卫生；软件

行业分类：

研究起止时间：

课题来源：

课题立项编号：

课题立项名称：

经费实际投入额：万元

应用状态：

转让范围：

推广形式：

转让条件：

转让内容：

评价方式：

评价单位：

评价日期：

评价证书号：

鉴定日期：

鉴定证书号：

鉴定委员会名单：

鉴定意见：

完成单位情况

第一完成单位：齐齐哈尔医学院

单位地址：黑龙江省齐齐哈尔市富拉尔基区和平路 249 号

邮政编码：161042

联系电话：0452-6731348

联系人姓名：

传真：0452-6712437

合作完成单位：

完成人：

成果简介

中药材真伪鉴别资料库收集了正品中药材原色图谱,同时也收集了该品种的常见混淆品的原色图谱,以及显微鉴别、快速鉴别、光谱与色谱的鉴别、检验标准等资料,非常适用于中药材检验工作人员的使用。设计的软件建立自动检索方式、所收录的中药材的原色图谱具有放大与缩小功能,还可实现全屏显示。收载的项目有形状鉴别、来源、显微鉴别、快速鉴别、光谱鉴别、常见混淆品鉴别、检验标准。共完成原色图片 1029 幅、178 种快速鉴别方法、可鉴别 467 种常用的中药材。

二、统计分析

(一) 学科分类统计

| 门类 | 学科 | 成果数 |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------|-------|
| 基础医学 | 医学生物化学、人体解剖学、医学细胞生物学、人体生理学、人体组织胚胎学、医学遗传学、放射医学、人体免疫学、医学寄生虫学、医学微生物学(包括医学病毒学等)、病理学、药理学、医学实验动物学、医学心理学、医学统计学、生物医学工程学、基础医学其他学科 | 6355 |
| 临床医学 | 临床诊断学、保健医学、理疗学、麻醉学、内科学、外科学、妇产科学、儿科学、眼科学、耳鼻咽喉科学、口腔医学、皮肤病学、性医学、神经病学、精神病学(包括精神卫生及行为医学等)、急诊医学、核医学、肿瘤学、护理学、临床医学其他学科 | 35274 |
| 预防医学与卫生学 | 营养学、毒理学、消毒学、流行病学、传染病学、媒介生物控制学、环境医学、职业病学、地方病学、社会医学、卫生检验学、食品卫生学、儿少卫生学、妇幼卫生学、环境卫生学、劳动卫生学、放射卫生学、卫生工程学、卫生经济学、优生学、健康教育学、卫生管理学、预防医学与卫生学其他学科 | 5903 |
| 中医学与中药学 | 中医学、民族医学、中西医结合医学、中药学、中医学与中药学其他学科 | 9599 |
| 军事医学与特种医学 | 军事医学、特种医学、军事医学与特种医学其他学科 | 168 |
| 药 学 | 药剂学、药物化学(包括天然药物化学等)、生物药物学、微生物药物学、放射性药物学、药效学、药物管理学、药物统计学、药学其他学科 | 5891 |

(二) 成果体现形式

表 1 成果体现形式

| 科技成果状态 | 类别 | 数量 |
|--------|----------|-------|
| 成果体现形式 | 新技术 | 14788 |
| | 新产品 | 2470 |
| | 新工艺 | 349 |
| | 农业、生物新品种 | 168 |
| 成果水平 | 国际领先 | 1472 |
| | 国际先进 | 5611 |
| | 国内领先 | 16100 |

| 科技成果状态 | 类别 | 数量 |
|----------|--------------|-------|
| 推广形式 | 国内先进 | 7209 |
| | 国内一般 | 1286 |
| | 产权转让 | 451 |
| | 资金入股 | 62 |
| | 技术入股 | 209 |
| | 合作开发 | 1090 |
| | 技术服务 | 4806 |
| 成果所处阶段 | 其他 | 19152 |
| | 成熟应用阶段 | 15580 |
| | 中期阶段 | 6935 |
| | 初期阶段 | 5396 |
| 国家计划成果分布 | 863 计划 | 270 |
| | 国家科技攻关计划 | 749 |
| | 基础研究计划 | 805 |
| | 研究开发条件能力建设 | 45 |
| | 科技产业化环境建设计划 | 36 |
| | 国家科技基础条件平台计划 | 4 |
| 近五年成果分布 | 2007 年 | 5222 |
| | 2006 年 | 6287 |
| | 2005 年 | 5349 |
| | 2004 年 | 4543 |
| | 2003 年 | 2557 |

（三）医药卫生领域内研发热点统计

以国家科技成果库中收录的 2000 年（含）以后的登记成果为主要依据进行统计分析，旨在反映各细分领域近年来的研发热点。

表 2 医药卫生领域内研发热点统计

| 热点领域 | 成果总数 | 研发单位数 | 三年内成果数量分布 | | |
|------|------|-------|-----------|--------|--------|
| | | | 2007 年 | 2006 年 | 2005 年 |
| 治疗 | 1261 | 693 | 391 | 340 | 138 |
| 糖尿病 | 428 | 286 | 38 | 96 | 101 |
| 诊断 | 344 | 245 | 60 | 61 | 45 |
| 儿童 | 321 | 219 | 38 | 76 | 52 |
| 冠心病 | 297 | 205 | 59 | 58 | 51 |

三、领域研发单位统计与分析

以国家科技成果库中收录的 2000 年（含）以后的登记成果为主要依据进行

统计分析，旨在反映各细分领域研发单位的研发重点、水平等方面情况，便于企业有针对性地选择科研单位，开展技术合作，解决技术难题。

表 3 该领域内科研机构业绩排名

| 排名 | 单位名称 | 单位属性 | 累计成果 数量 | 计划项目 成果数量 | 专利数 |
|----|-------------------|------|------------|--------------|-----|
| 1 | 天津医科大学总医院 | 医疗机构 | 213 | 50 | 15 |
| 2 | 潍坊市人民医院 | 医疗机构 | 189 | 168 | 2 |
| 3 | 华中科技大学同济医学院 | 大专院校 | 162 | 55 | 2 |
| 4 | 河北医科大学第二医院 | 医疗机构 | 152 | 61 | 0 |
| 5 | 临沂市人民医院 | 医疗机构 | 151 | 148 | 0 |
| 6 | 上海交通大学医学院附属第九人民医院 | 医疗机构 | 150 | 74 | 26 |
| 7 | 河北省人民医院 | 医疗机构 | 144 | 71 | 0 |
| 8 | 滨州医学院附属医院 | 医疗机构 | 142 | 3 | 2 |
| 9 | 天津市第一中心医院 | 医疗机构 | 141 | 15 | 9 |
| 10 | 昆明医学院第一附属医院/云大医院 | 医疗机构 | 140 | 23 | 2 |

附录 4 通过 Web of Science 查询在 SCI 期刊上发表肿瘤研究的科技论文

一、检索方法

检索式：主题词：cancer and 地址：China and 类型：article

二、检索结果

1、一般情况

ISI Web of KnowledgeSM 领先一步

所有数据库 选择一个数据库 Web of Science 其他资源

检索 被引参考文献检索 高级检索 检索历史 标记结果列表 (0)

Web of Science®

<< 返回前一结果

检索结果 Topic=(cancer) AND Address=(china)
入库时间=所有年份. 数据库=SCI-EXPANDED.
精炼依据: Document Type=(ARTICLE)

检索结果: **15,655** 页 1 /1,566 转

打印 电子邮件 添加到标记结果列表 保存到 EndNote Web

精炼检索结果

结果内检索

学科类别 精炼

- ☐ ONCOLOGY (4,916)
- ☐ BIOCHEMISTRY & MOLECULAR BIOLOGY (2,107)
- ☐ GASTROENTEROLOGY & HEPATOLOGY (1,466)
- ☐ PHARMACOLOGY & PHARMACY (1,409)
- ☐ CELL BIOLOGY (925)
- [更多选项/分类...](#)

文献类型 精炼

- ☐ ARTICLE (15,655)

作者

来源出版物

出版年

机构

语种

国家/地区

分析检索结果

1. Title: E-Cadherin Tissue Expression and Urinary Soluble Forms
Author(s): Shi BK, Laudon V, Yu SQ, et al.
Source: UROLOGIA INTERNATIONALIS Volume: 81 Issue: 3 Pages: 1-4
Times Cited: 0
[Check For Full Text](#)






















2. Title: Rapid Membrane Effect of Testosterone in LNCaP Cells
Author(s): Wang ZF, Liu LM, Hou JQ, et al.
Source: UROLOGIA INTERNATIONALIS Volume: 81 Issue: 3 Pages: 1-4
Times Cited: 0
[Check For Full Text](#)

3. Title: OPTIMIZATION OF ULTRASOUND-MEDIATED IN VITRO LINE HEPG2
Author(s): Zhai BJ, Shao ZY, Zhao CL, et al.
Source: ULTRASOUND IN MEDICINE AND BIOLOGY Volume: 34 Issue: 3 Pages: 1-4
Times Cited: 0
[Check For Full Text](#)

4. Title: Long-term accumulation and low toxicity of single-walled carbon nanotubes in mice
Author(s): Yang ST, Wang X, Jia G, et al.
Source: TOXICOLOGY LETTERS Volume: 181 Issue: 3 Pages: 1-4
Times Cited: 0
[Check For Full Text](#)

5. Title: Genistein induces cell apoptosis in MDA-MB-231 breast cancer cells
Author(s): Li Z, Li J, Mo BQ, et al.
Source: TOXICOLOGY IN VITRO Volume: 22 Issue: 7 Pages: 17-24
Times Cited: 0


































2、学科分布

| 字段:学科类别 | 记录数 | % , 共15655 | 柱状图 |
|-----------------------------------------------|------|------------|---------------------------------------------------------------------------------------|
| ONCOLOGY | 4916 | 31.4021 % |  |
| BIOCHEMISTRY & MOLECULAR BIOLOGY | 2107 | 13.4590 % |  |
| GASTROENTEROLOGY & HEPATOLOGY | 1456 | 9.3005 % |  |
| PHARMACOLOGY & PHARMACY | 1409 | 9.0003 % |  |
| CELL BIOLOGY | 925 | 5.9087 % |  |
| BIOPHYSICS | 842 | 5.3785 % |  |
| MEDICINE, RESEARCH & EXPERIMENTAL | 725 | 4.6311 % |  |
| CHEMISTRY, MULTIDISCIPLINARY | 640 | 4.0882 % |  |
| GENETICS & HEREDITY | 631 | 4.0307 % |  |
| MEDICINE, GENERAL & INTERNAL | 597 | 3.8135 % |  |
| SURGERY | 553 | 3.5324 % |  |
| CHEMISTRY, MEDICINAL | 541 | 3.4558 % |  |
| PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH | 479 | 3.0597 % |  |
| RADIOLOGY, NUCLEAR MEDICINE & MEDICAL IMAGING | 474 | 3.0278 % |  |
| BIOTECHNOLOGY & APPLIED MICROBIOLOGY | 463 | 2.9575 % |  |
| BIOCHEMICAL RESEARCH METHODS | 383 | 2.4465 % |  |
| PATHOLOGY | 361 | 2.3060 % |  |
| TOXICOLOGY | 351 | 2.2421 % |  |
| IMMUNOLOGY | 322 | 2.0569 % |  |
| CHEMISTRY, ANALYTICAL | 320 | 2.0441 % |  |
| OBSTETRICS & GYNECOLOGY | 311 | 1.9866 % |  |
| CHEMISTRY, ORGANIC | 294 | 1.8780 % |  |
| ENDOCRINOLOGY & METABOLISM | 237 | 1.5139 % |  |
| CHEMISTRY, APPLIED | 236 | 1.5075 % |  |
| UROLOGY & NEPHROLOGY | 222 | 1.4181 % |  |
| PLANT SCIENCES | 206 | 1.3159 % |  |
| ENVIRONMENTAL SCIENCES | 194 | 1.2392 % |  |
| HEMATOLOGY | 186 | 1.1881 % |  |
| FOOD SCIENCE & TECHNOLOGY | 185 | 1.1817 % |  |
| BIOLOGY | 176 | 1.1242 % |  |
| RESPIRATORY SYSTEM | 156 | 0.9965 % |  |
| MULTIDISCIPLINARY SCIENCES | 152 | 0.9709 % |  |
| MEDICAL LABORATORY TECHNOLOGY | 131 | 0.8368 % |  |

3、来源期刊分布

| 字段:来源出版物 | 记录数 | %，共15655 | 柱状图 |
|-------------------------------------------------------------|-----|----------|-----|
| WORLD JOURNAL OF GASTROENTEROLOGY | 940 | 6.0045 % | ■ |
| CHINESE MEDICAL JOURNAL | 445 | 2.8425 % | ■ |
| INTERNATIONAL JOURNAL OF CANCER | 314 | 2.0057 % | ■ |
| CANCER RESEARCH | 295 | 1.8844 % | ■ |
| CANCER LETTERS | 271 | 1.7311 % | ■ |
| CLINICAL CANCER RESEARCH | 258 | 1.6480 % | ■ |
| BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS | 254 | 1.6225 % | ■ |
| ONCOLOGY REPORTS | 211 | 1.3478 % | ■ |
| PROGRESS IN BIOCHEMISTRY AND BIOPHYSICS | 205 | 1.3095 % | ■ |
| ANTICANCER RESEARCH | 177 | 1.1306 % | ■ |
| ACTA PHARMACOLOGICA SINICA | 166 | 1.0604 % | ■ |
| CANCER BIOLOGY & THERAPY | 163 | 1.0412 % | ■ |
| CARCINOGENESIS | 163 | 1.0412 % | ■ |
| JOURNAL OF CANCER RESEARCH AND CLINICAL ONCOLOGY | 161 | 1.0284 % | ■ |
| ACTA BIOCHIMICA ET BIOPHYSICA SINICA | 143 | 0.9134 % | ■ |
| CANCER EPIDEMIOLOGY BIOMARKERS & PREVENTION | 136 | 0.8687 % | ■ |
| ONCOGENE | 136 | 0.8687 % | ■ |
| CANCER | 127 | 0.8112 % | ■ |
| INTERNATIONAL JOURNAL OF ONCOLOGY | 111 | 0.7090 % | ■ |
| GYNECOLOGIC ONCOLOGY | 108 | 0.6899 % | ■ |
| BRITISH JOURNAL OF CANCER | 104 | 0.6643 % | ■ |
| JOURNAL OF BIOLOGICAL CHEMISTRY | 101 | 0.6452 % | ■ |
| BMC CANCER | 86 | 0.5493 % | ■ |
| INTERNATIONAL JOURNAL OF GYNECOLOGICAL CANCER | 84 | 0.5366 % | ■ |
| INTERNATIONAL JOURNAL OF MOLECULAR MEDICINE | 82 | 0.5238 % | ■ |
| LIFE SCIENCES | 82 | 0.5238 % | ■ |
| CHINESE JOURNAL OF CANCER RESEARCH | 78 | 0.4982 % | ■ |
| SPECTROSCOPY AND SPECTRAL ANALYSIS | 77 | 0.4919 % | ■ |
| LUNG CANCER | 76 | 0.4855 % | ■ |
| BIOORGANIC & MEDICINAL CHEMISTRY LETTERS | 75 | 0.4791 % | ■ |
| CHINESE SCIENCE BULLETIN | 74 | 0.4727 % | ■ |
| INTERNATIONAL JOURNAL OF RADIATION ONCOLOGY BIOLOGY PHYSICS | 72 | 0.4599 % | ■ |

4、机构分布

| 字段:机构名称 | 记录数 | % , 共15655 | 柱状图 |
|-----------------------------|------|------------|---------------------------------------------------------------------------------------|
| UNIV HONG KONG | 1156 | 7.3842 % |  |
| CHINESE ACAD SCI | 1131 | 7.2245 % |  |
| CHINESE UNIV HONG KONG | 941 | 6.0109 % |  |
| FUDAN UNIV | 805 | 5.1421 % |  |
| PEKING UNIV | 763 | 4.8738 % |  |
| CHINESE ACAD MED SCI | 744 | 4.7525 % |  |
| ZHEJIANG UNIV | 713 | 4.5545 % |  |
| CHINA MED UNIV | 504 | 3.2194 % |  |
| SUN YAT SEN UNIV | 479 | 3.0597 % |  |
| PEKING UNION MED COLL | 428 | 2.7340 % |  |
| SHANGHAI JIAO TONG UNIV | 416 | 2.6573 % |  |
| FOURTH MIL MED UNIV | 413 | 2.6381 % |  |
| HUAZHONG UNIV SCI & TECHNOL | 412 | 2.6317 % |  |
| SHANDONG UNIV | 390 | 2.4912 % |  |
| SICHUAN UNIV | 342 | 2.1846 % |  |
| WUHAN UNIV | 315 | 2.0121 % |  |
| NCI | 305 | 1.9483 % |  |
| MIL MED COLL 2 | 262 | 1.6736 % |  |
| UNIV TEXAS | 257 | 1.6416 % |  |
| SHANGHAI CANC INST | 247 | 1.5778 % |  |
| NANJING MED UNIV | 237 | 1.5139 % |  |
| NANJING UNIV | 216 | 1.3798 % |  |
| CHINA MED UNIV HOSP | 209 | 1.3350 % |  |
| CHINA MED COLL HOSP | 198 | 1.2648 % |  |
| XIAN JIAOTONG UNIV | 191 | 1.2201 % |  |
| SHENYANG PHARMACEUT UNIV | 175 | 1.1179 % |  |
| CENT S UNIV | 168 | 1.0731 % |  |
| VANDERBILT UNIV | 165 | 1.0540 % |  |
| TIANJIN MED UNIV | 163 | 1.0412 % |  |
| CHINA MED COLL | 162 | 1.0348 % |  |
| HONG KONG POLYTECH UNIV | 160 | 1.0220 % |  |
| THIRD MIL MED UNIV | 157 | 1.0029 % |  |
| HARBIN MED COLL | 154 | 0.9837 % |  |

5、年代分布

| 字段:出版年 | 记录数 | % , 共 15655 | 柱状图 |
|--------|------|-------------|-----|
| 1996 | 10 | 0.0639 % | |
| 1997 | 199 | 1.2712 % | |
| 1998 | 265 | 1.6927 % | |
| 1999 | 339 | 2.1654 % | |
| 2000 | 525 | 3.3536 % | |
| 2001 | 694 | 4.4331 % | |
| 2002 | 882 | 5.6340 % | |
| 2003 | 1203 | 7.6844 % | |
| 2004 | 1272 | 8.1252 % | |
| 2005 | 1628 | 10.3992 % | |
| 2006 | 2315 | 14.7876 % | |
| 2007 | 3163 | 20.2044 % | |
| 2008 | 3159 | 20.1789 % | |

附录 5 万方科技成果数据库查询和分析结果

一、调查内容

- 1、调查范围：中国科技成果库、中国重大科技成果库、科技成果交易库、科技成果精品数据库、国家级科技授奖项目数据库、科技决策支持库等 7 个成果库。
- 2、查询策略

二、查询结果

（一）基本情况

表 1 肿瘤领域科技成果类情况

| 类别 | 数量 |
|------------|--------|
| 国家级科技授奖成果 | 12 |
| 科技决策支持成果 | 4 |
| 中国重大科技成果 | 339 |
| 中国科技成果 | 3, 229 |
| 全国科技成果交易信息 | 926 |

附录 6 肿瘤主题数据库框架设计结果

一、肿瘤科学数据关联的业务活动分析结果

1、肿瘤的预防控制

肿瘤的预防控制主要是基于人群或群体的肿瘤的流行病学研究和肿瘤疾病的预防控制活动。具体工作包括调查肿瘤疾病在人群中分布和影响分布的因素，病因探索、制定防制肿瘤的策略和措施、评价实施效果等。所涉及的主题信息类包括社会因素、自然环境因素、心理和行为、肿瘤的发生和死亡、人群特征等。其业务活动大致包括：

- 一肿瘤有关的健康调查与监测

- 一肿瘤环境调查与监测

- 一癌症综合防治

- 一肿瘤的专题研究

肿瘤的预防控制数据多数为人群调查记录、实验室检测、和公共卫生服务业务数据以及在此基础上产生的分析和报告数据等。

2、肿瘤基础研究

肿瘤基础研究从分子、细胞、组织、器官、机体、遗传等层面对肿瘤的发生和发展规律揭示。例如从分子生物学的角度，恶性肿瘤可视为基因的疾病，是某些染色体上的 DNA 损伤致使基因突变的结果，导致细胞的生长失控件、缺乏分化而异常增生，并可侵犯正常组织和器官，最终可散布全身。从细胞学角度，肿瘤是由基因突变中导致异常增生的单个细多阶段胞，这种异常增生细胞克隆出来的后裔所形成。

肿瘤基础研究涉及到与多个学科的交叉，一般来说更微观层次类特征往往是高层次对象变化的原因和解释（垂直推理现象）。例如肿瘤的克隆性起源并不意味着产生肿瘤的原始细胞从一开始就已获得了恶性细胞的所有特征。相反，恶性肿瘤的发生是一个逐步演变的过程，肿瘤细胞是通过一系列进行性的改变而逐渐变成肤件的。在这种克隆性演化过程中，常积累一系列的基因突变，可涉及不问染色体多种基因的变化，包括：癌基因、肿瘤抑制基因、细胞周期调节基因、细胞凋亡基因及维持细胞基因组稳定性的基因(包括：DNA 修复、DNA 复制及染色体分离基因)等。同样，肿瘤病理变化需要通过研究肿瘤细胞的特点才能得以理解，

与肿瘤细胞的生长和凋亡离不开对肿瘤基因的分析一样。

对于肿瘤基础研究，科学家的研究活动的目标往往相对固定在某个层面上，其研究活动、以及相伴产生数据和结果往往用于描述或解释某个层面上的科学问题。因此根据研究数据所需要解决问题层面性进行分类可解决数据集多主题或多层次主题交叉的问题。

肿瘤基础研究多来自实验室，对于相同对象采用不同的实验技术其原始数据的也相差较大。

3、肿瘤临床研究

肿瘤临床研究主要是针对个体的，通过诊断、治疗、评价等研究或业务活动，达到消灭肿瘤、促进康复和延长寿命的目的。在一定程度上来说是基础研究的应用研究。临床诊断活动通过不同的诊断技术对与肿瘤不同层次的对象不同属性进行测量和描述，如实验室生化诊断、影像、病理等数据；临床治疗根据所采用的治疗手段或工具及其适应性的不同而产生不同结果数据。对临床治疗的效果评价涉及到并发症、康复效果、以及生存时间等问题。

4、抗肿瘤药物

其业务活动包括产生的各种数据的抗肿瘤药物研发过程，如化学成分鉴定、药理实验、药物毒理实验、药物临床实验、副作用监测、药物登记和流通管理等。

二、肿瘤主题数据库资源框架

| 一级类目 | 二级类目 | 三级类目 |
|---------|-------|-----------|
| 肿瘤预防与控制 | 肿瘤监测 | 肿瘤死因监测 |
| | | 肿瘤登记和发病监测 |
| | | 危险因素监测 |
| | 肿瘤调查 | 肿瘤患病和发病调查 |
| | | 肿瘤死因调查 |
| | | 肿瘤危险因素调查 |
| | | 卫生学调查 |
| | 检测和检查 | 卫生学检测 |
| | | 健康检查 |
| | | 肿瘤筛查 |

| 一级类目 | 二级类目 | 三级类目 |
|--------|-----------|----------------------------------------------------------------------------------------------------|
| 肿瘤基础研究 | 肿瘤干预 | 卫生政策干预 危险因素干预 营养和药物干预 |
| | 肿瘤预防和控制评价 | 健康状况评价 干预效果评价 卫生状况评价 |
| | 报告和发布 | 统计报告 调查结果报告 干预结果报告 |
| | 专题研究 | |
| | 分子水平机制研究 | 核苷酸序列数据 核酸序列数据 蛋白质序列数据 结构数据 代谢和信号转导途径 其他分子生物学数据 肿瘤基因库 微阵列数据和其他基因表达 蛋白质组学资源 |
| | 细胞水平机制研究 | 细胞结构数据 细胞功能数据 肿瘤干细胞研究 |
| | 肿瘤组织病理研究 | 病理组织形态 病理组织化学 |
| | 肿瘤免疫机制研究 | 细胞免疫 体液免疫 免疫疫苗 |
| | 实验标本和模型 | 组织培养 移植与接种 病理组织库资源 模式生物库 相关试验材料 |
| | | |
| 肿瘤临床研究 | 肿瘤诊断研究 | |

| 一级类目 | 二级类目 | 三级类目 |
|------------|--------------|----------|
| 抗肿瘤药物管理和研究 | 肿瘤临床治疗研究 | 实验室诊断 |
| | | 肿瘤的病理诊断 |
| | | 影像学诊断 |
| | | 肿瘤的免疫学诊断 |
| | | 肿瘤的内腔镜诊断 |
| | | 超声波诊断 |
| | | 化学（药物）治疗 |
| | | 放射治疗 |
| | | 手术（外科）治疗 |
| | | 免疫治疗 |
| | 肿瘤临床转归研究 | 肿瘤基因治疗 |
| | | 中医药治疗 |
| | | 激光治疗 |
| | | 综合治疗 |
| | | 姑息治疗和其他 |
| | | 肿瘤并发症研究 |
| | | 肿瘤预后研究 |
| | | 肿瘤康复研究 |
| | | 药物研究数据 |
| 药理实验 | | |
| 药物毒理实验 | | |
| 药物临床实验 | | |
| 副作用监测 | | |
| 药物管理数据 | 抗肿瘤药物药品登记 | |
| 通用数据 | 术语和标准 | |
| | 人口、社会和环境基础信息 | |
| | 实验和诊断试剂 | |
| | 实验和诊断器材和器械信息 | |

附录 7 资源信息管理工具/系统

一、基本功能

提供资源信息数据库中的数据查询、编辑、排序、以及删除和添加功能。

二、数据模式

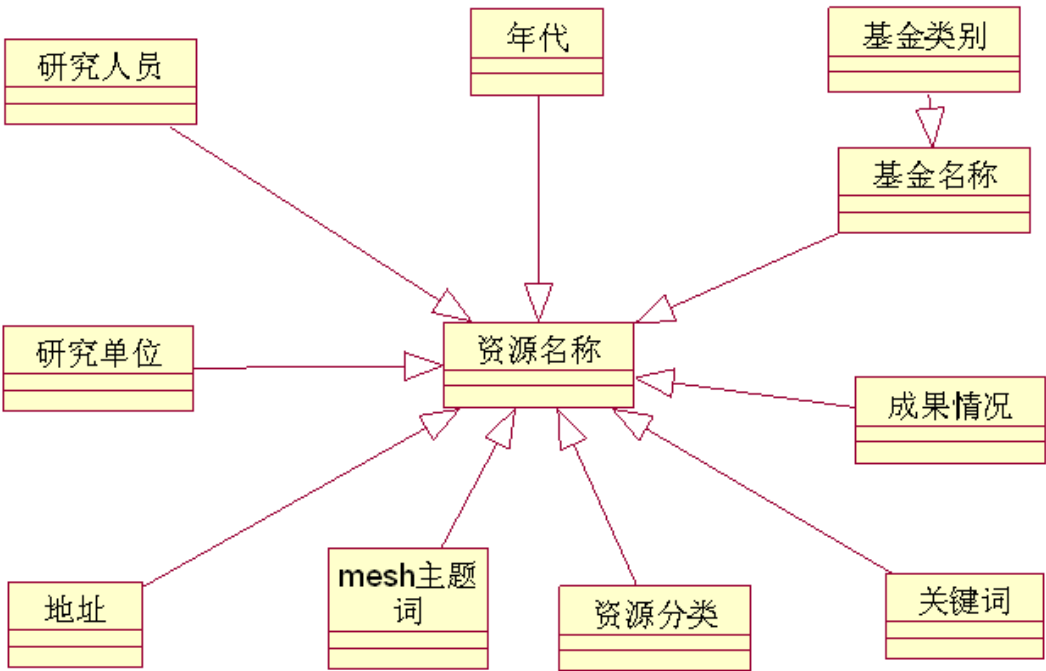


图 1 资源信息库核心数据元

三、运行环境

- 1、基于 PHP+MySQL 数据库开发；
- 2、需要配置 Apache+MySQL+PHP 的整体环境；在 Apache-2.0.63、PHP-5.2.6、MySQL-5.0.51b 的环境下通过测试。

四、基本界面介绍

暂略