

<<科学数据库与信息技术论文集>>

图书基本信息

书名：<<科学数据库与信息技术论文集>>

13位ISBN编号：9787030332424

10位ISBN编号：7030332423

出版时间：2012-1

出版时间：科学出版社

作者：中国科学院科学数据库办公室 编

页数：475

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

## <<科学数据库与信息技术论文集>>

### 内容概要

《科学数据库与信息技术论文集（第11集）》收入论文59篇，主要反映中国科学院近年来在科学数据库的建库技术、网络技术、信息服务技术等方面所取得的成果及学术上取得的进展。这些论文也集中体现了近年来国内数据库与信息技术方面的研究和应用水平。

《科学数据库与信息技术论文集（第11集）》可供从事数据库技术、网络技术和信息系统研究的科技人员、工程技术人员参考，也可供相关学科的研究人员、大专院校师生阅读。

## 书籍目录

序总论中国科学院科技数据资源现状及其发展思考科学数据的个体识别和跨学科集成数据库系统及建库技术化合物子结构检索及时响应的流程优化与分布式优化研究植物化学成分数据库建设中国植物物种信息数据库关联性设计与查询服务数字化物种编目的三个重要问题植物药用知识传承与利用的数据库建设与应用探索毫米波射电天文数据库研究化合物结构的唯一性编码和化学数据库管理空间数据存储技术及四大空间数据库发展浅析NoSQL当前发展及应用状况系统生物学中多组学综合数据库的设计与实现Web数据库安全技术分析及其应用LinkedData在科学数据库中的应用探讨数字版权保护与数字水印技术在科学数据库中的应用基于网络服务的开放矢量地理空间数据管理系统数据共享与服务技术人地系统科学数据网格建设初探化学主题数据库：数据网格技术的应用范例材料科学主题数据库的数据整合与共享服务空间科学数据网格模型应用服务框架设计与实现中国动物主题数据库服务平台的设计与实现卫星轨道演示软件的设计与实现面向数据密集型计算的科学数据网格作业调度系统一个轻量级分布式计算框架的设计与实现化合物理化性质数据的整合和服务基团贡献法物性估算模块的设计开发基于专业日志分析的科学数据库平台用户活动研究——以中国数字植物标本馆(CVH)为例基于Flex的基因组可视化工具基因组可视化工具GBrowse及其应用古生物学和地层学专业数据库中地层数据的可视化基于MRS工具的海量数据检索L,AS在南海物理海洋数据服务中的应用研究PHP常见安全漏洞研究基于HITS的科学数据检索结果排序的研究RDF在科学数据集成检索中的应用LEQL：一种面向科学数据实体的级联关联查询语言配置管理数据库(CMDB)的关键技术研究及实现大规模遥感图像高性能集群存储处理技术研究及实现科学数据管理与共享工具及技术研究通用数据源连接器设计及实现973计划资源环境领域数据汇交服务实践与成效移动互联网环境下群组日程协同软件设计与实现英汉冰冻圈科学词汇收录修订与共享系统数据共享与标准规范文献数据训及全文检索技术数据库应用

## 章节摘录

版权页：插图：理论上，一个有效的QSAR/QSPR模型可以预测得到化合物的多种性质，可以用于对化合物数据库的批量筛选。

但在实际建立与应用QSAR / QSPR模型时，常常会出现一些错误，从而导致预测得到的结果不合理。文献总结了在建立QSA/QSPR模型过程中可能出现的20类错误，大体可分为数据集的质量、描述符的选择、模型的验证、模型应用范围的定义、机理解释等几个方面。

为使QSAR / QSPR模型尽可能合理，在建立模型时，要尽量保证研究对象数据集的完备性、多样性、合理性、来源一致性。

在选择描述符时，要尽量选择与作用机理可能相关、具有明确物理意义的描述符，要保证描述符之间不相关。

模型建立之后，要对模型进行内部和外部验证，以检验模型的拟合可优度、鲁棒性和预测能力。

另外，由于QSAR / QSPR模型只对与训练集相似的化合物可以得到好的预测结果，模型应用范围会直接影响QSAR / QSPR模型的预测可靠性，因此QSAR/QSPR模型应用范围的定义也是非常重要的。

目前定义模型应用域范围的方法主要分为以下四类。

(1) 以片段为基础的方法。

如果训练集的片段集中不包含待测化合物的组成片段，将定义该待测化合物在应用域范围外。

对于该方法，将化合物拆分为片段的算法会影响应用域范围，片段包含化学信息越多，应用域范围越严格。

(2) 化学描述符的值范围方法。

用多个化学描述符组成的 维空间定义应用域。

但是有些描述符不能对应用域内外的化合物进行很好的区分，如文献[19]采用logP对测试集进行区分，几乎整个测试集都在应用域范围内，而对同样的数据集采用其他方法区分，其中一部分化合物在应用域范围外。

(3) 结构相似性方法。

根据相似的结构具有相似的化学性质这一假设，采用结构相似性方法应是更加合理的，但结构相似性是一个主观概念，判断结构相似的方法不同，可能应用域范围也不同；另外即使结构相似的化合物，也可能由于其作用机理不一样，导致性质不一样。

(4) 组合方法。

文献[19]表明以片段为基础的方法和结构相似性方法结合是一种更好的定义应用域范围的方法。

模型应用范围的定义是比较困难的，一方面化合物结构相似的判断具有主观性；另一方面有些化合物尽管结构相似，由于作用机理不同，导致性质产生很大差异；其次是所定义的模型应用域的方法要能够通过编程实现。

目前模型应用域定义仍然是QSAR/QSPR的研究热点，人们仍然在探索中。

一个合理的QSAR/QSPR模型，应该尽可能给出模型中的描述符与预测性质之间机理相关的解释。

需要注意的是，QSAR/QSPR模型包含的描述符只能说明与其预测性质相关，它们之间并不一定存在因果关系，因此模型包含的一些描述符号可能与作用机理是无关的，这也是导致一部分QSAR/QSPR模型预测能力低的一个原因。

如果在建立模型之前，作用机理已经明确，根据作用机理建立的模型，其预测能力应该会大大提高。

编辑推荐

《科学数据库与信息技术论文集(第11集)》由科学出版社出版。

版权说明

本站所提供下载的PDF图书仅提供预览和简介, 请支持正版图书。

更多资源请访问:<http://www.tushu007.com>