

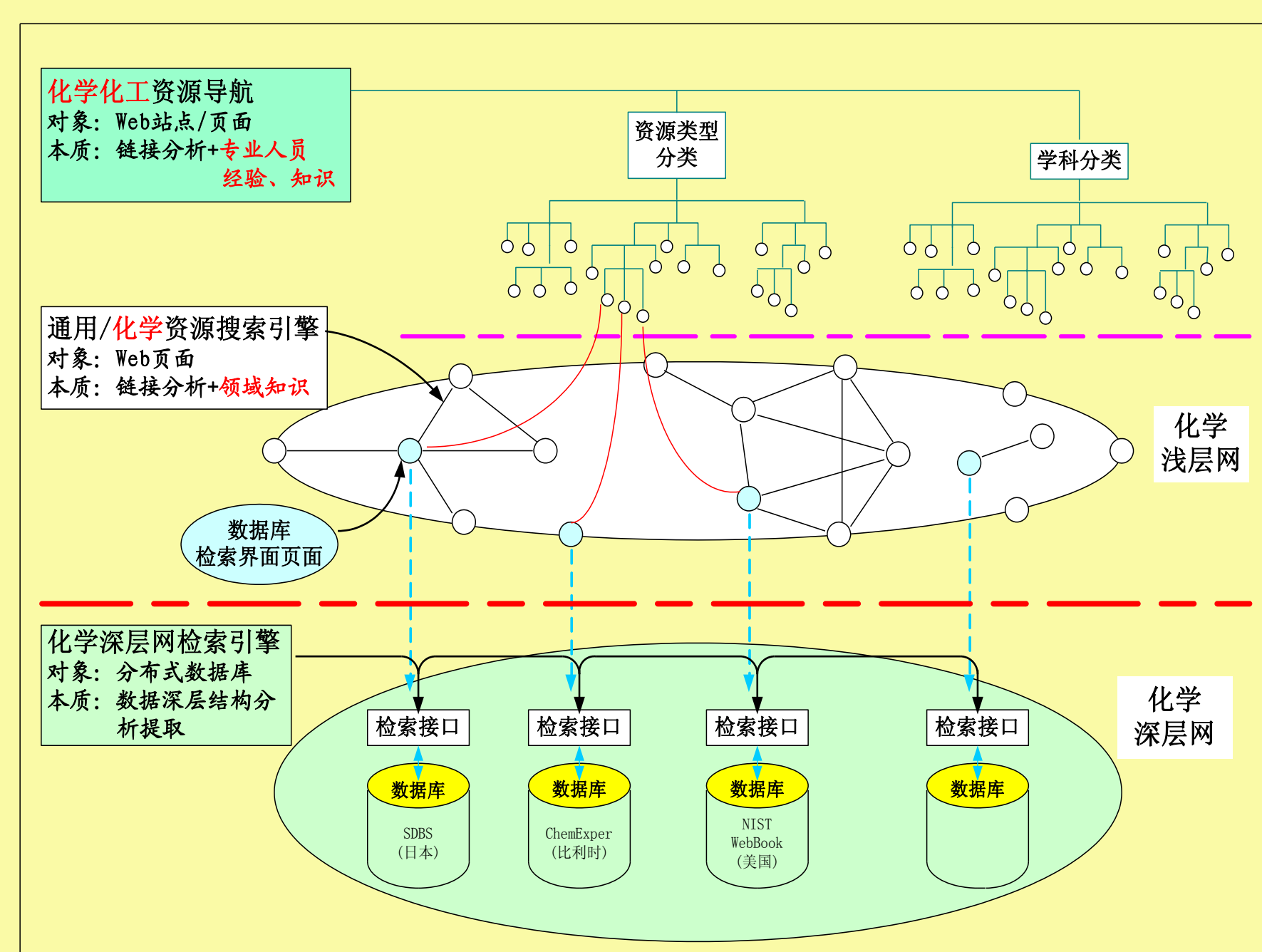


高性能计算与网络化学化工专业信息检索

高性能计算与化学信息学研究组
中国科学院过程工程研究所



网络化学化工专业信息检索



化学Web资源检索工具	专业资源导航 ChIN	专业搜索引擎 ChemEngine	深层网挖掘 ChemDB Portal
对象 (网络化学资源)	Web站点/页面	Web页面	分布式数据库中的数据
本质 (信息学+领域知识)	链接分析 + 信息源知识 + 领域知识	链接分析 + 领域知识	数据深层结构分析 + 领域知识
方法特点	人工参与资源的发现、描述、组织	智能程序; 控制爬行、自动分类	自动提交、数据提取、数据集成
主要技术	DBMS+Web交互	Web Robot + 文本自动分类 + 全文检索	XML+网络+?
专门工具	ChIN-Manager 资源描述与自动发布	控制爬行器、索引器、分类器、检索系统	数据提取模板 XSL生成工具+?
系统规模	> 10681 Web站点/页面	> 1000万 Web页面	9个化学数据库
成熟度	国内权威 国际承认	初步测试优于现有的两个同类系统	未见同类系统 正在研制

该方向的工作已超过10年, 研究逐步深入并系统化, 形成了系列检索工具。

多相复杂系统中的高性能科学计算



针对多相复杂系统, 依托独特的计算机集群硬件平台, 综合采用高性能并行计算技术, 对粒子方法模拟的各个环节提出优化算法, 以解决粒子方法模拟计算量巨大的瓶颈问题。

新的集群硬件方案

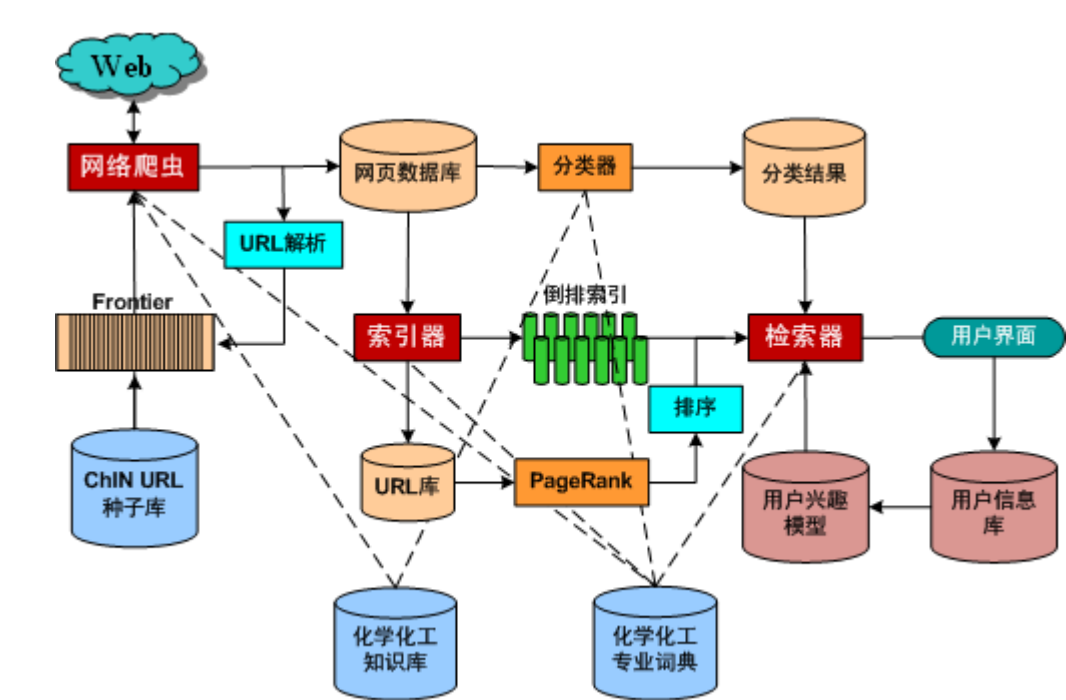
高性价比Tree+Mesh构型、粒子方法并行模拟专用高性能计算集群的可行性研究与系统设置。

数据前处理

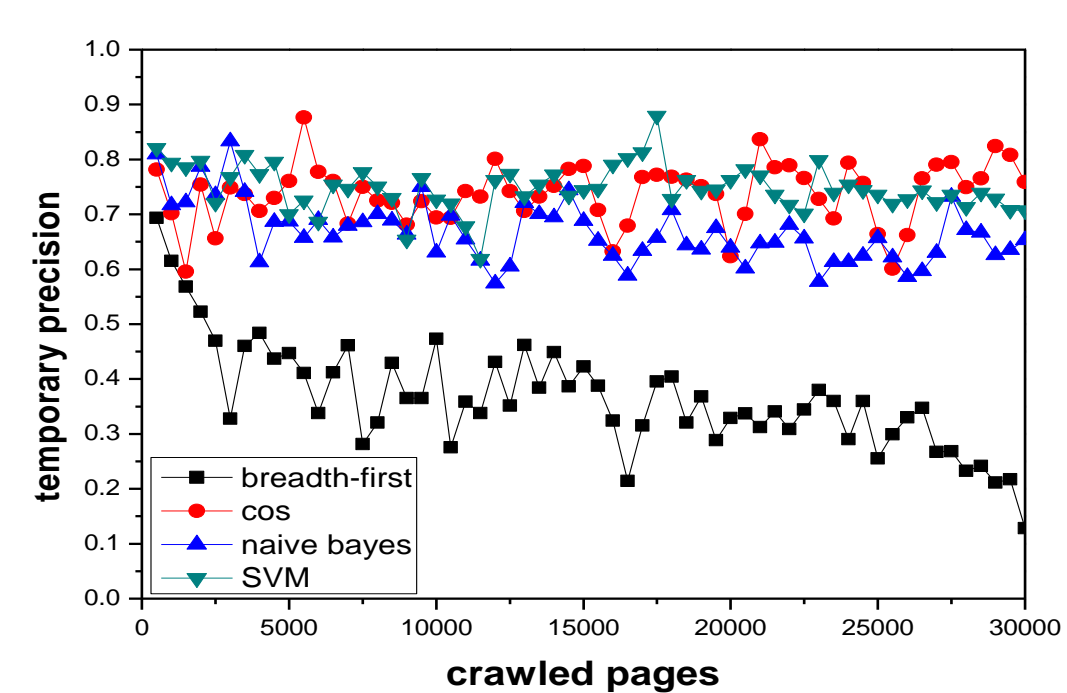
采用集合实体造型 (CSG) 技术实现的3D交互式模拟区域实体建模系统与粒子数据预处理系统。

网络化学化工专业搜索引擎

将化学化工领域知识与搜索引擎技术相结合, 采用机器学习的方法, 通过文本自动分类和专业信息选择性爬行, 建立了规模为1000万页的化学化工专业搜索引擎ChemEngine。



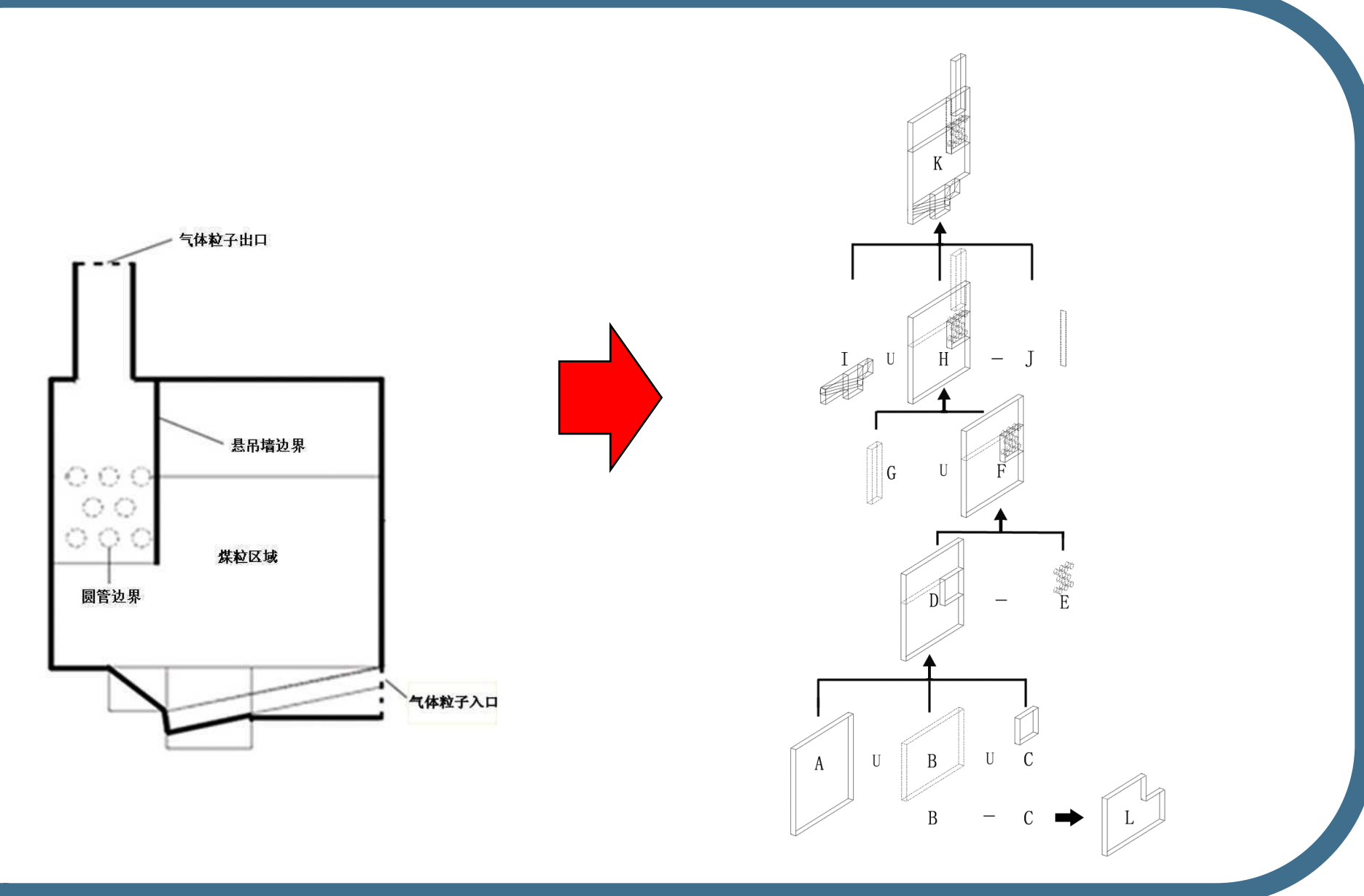
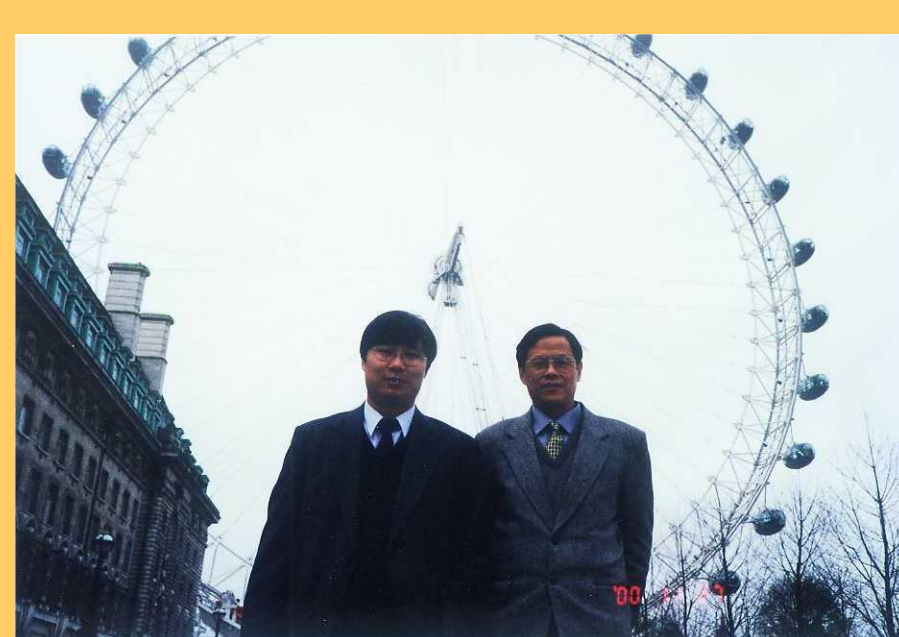
ChemEngine系统框架



ChemEngine主题爬虫爬行策略与通用搜索引擎比较



本课题组研究方向独特、工作系统队伍精干、敬业、富有职业精神



某煤炉燃烧模拟的3D建模过程

编程模式

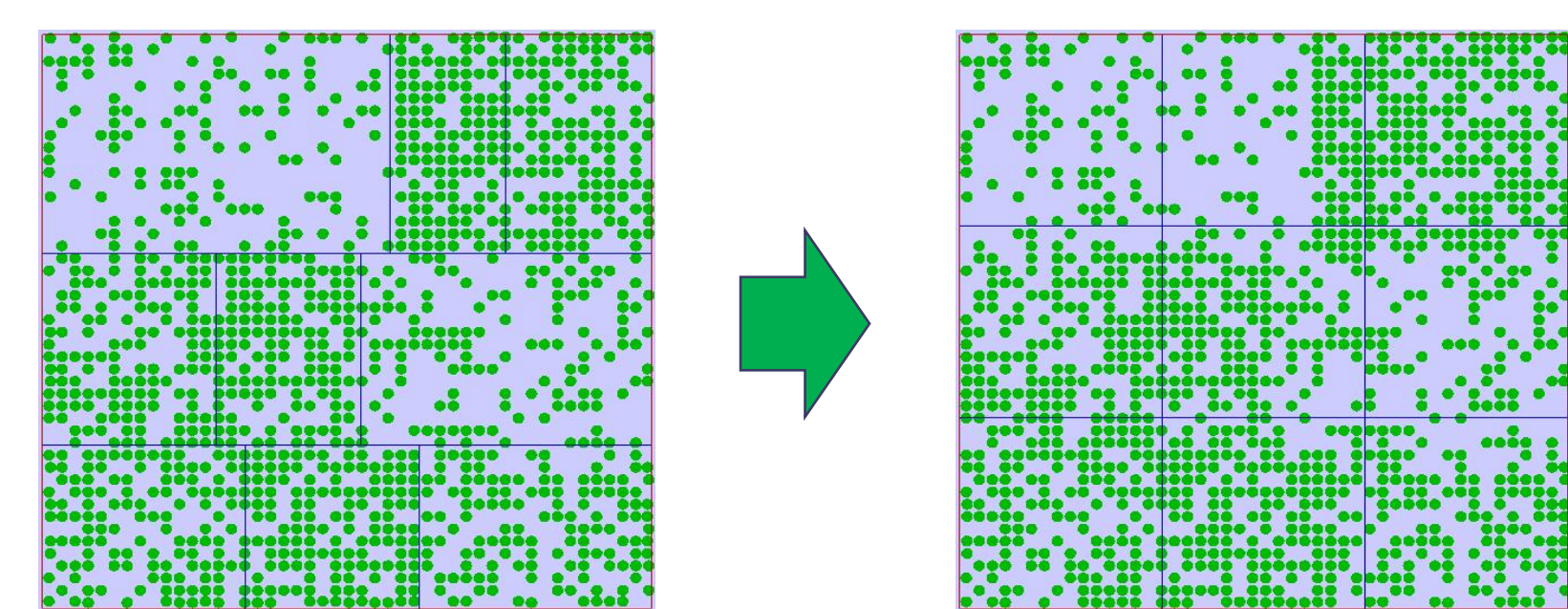
针对多CPU、多核、内存共享节点构成的高性能计算集群采用MPI+OpenMP编程模式和软件流水线技术提高模拟计算效率。

All2All通信模式

设计实现了单一类型粒子模拟的All2All并行通信模式和多类型粒子代码自动生成的泛型化, 克服了Shift通信模式的局限性。

计算负载均衡

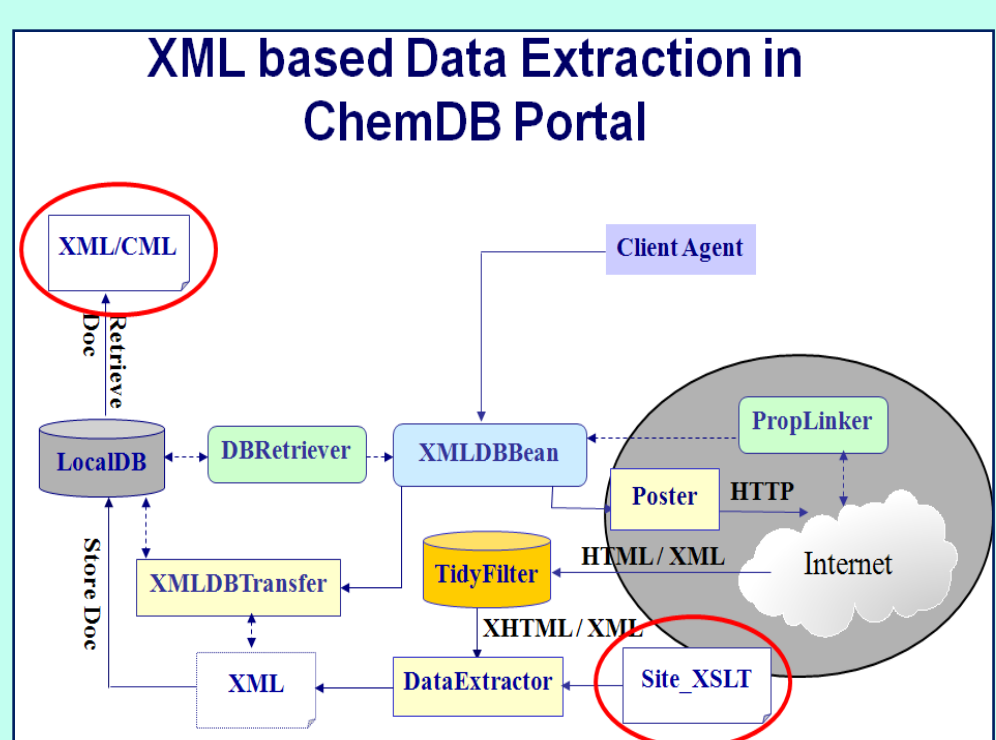
基于Master/Slave模式, 设计集群节点计算负载定时监控系统, 在负载不均衡时对计算区域进行自动重新划分。



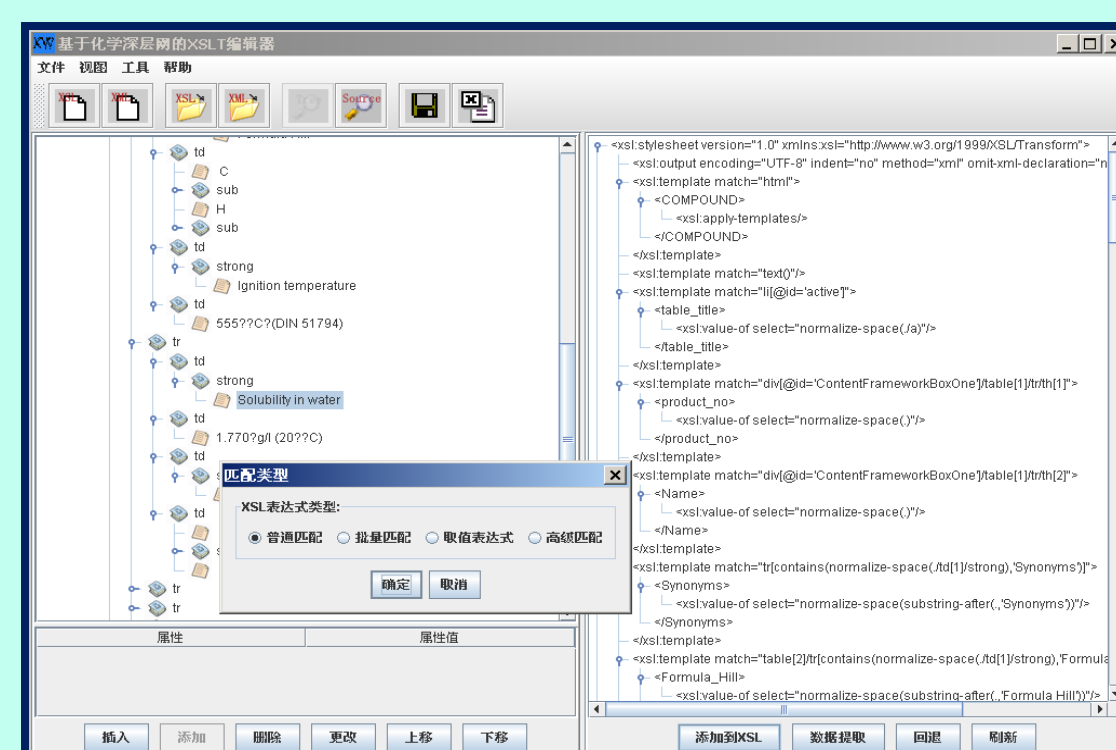
静态划分 (计算负载不均衡) 动态划分 (计算负载均衡)

稀薄气体分子动力学并行模拟中区域的计算负载静态划分和动态划分比较

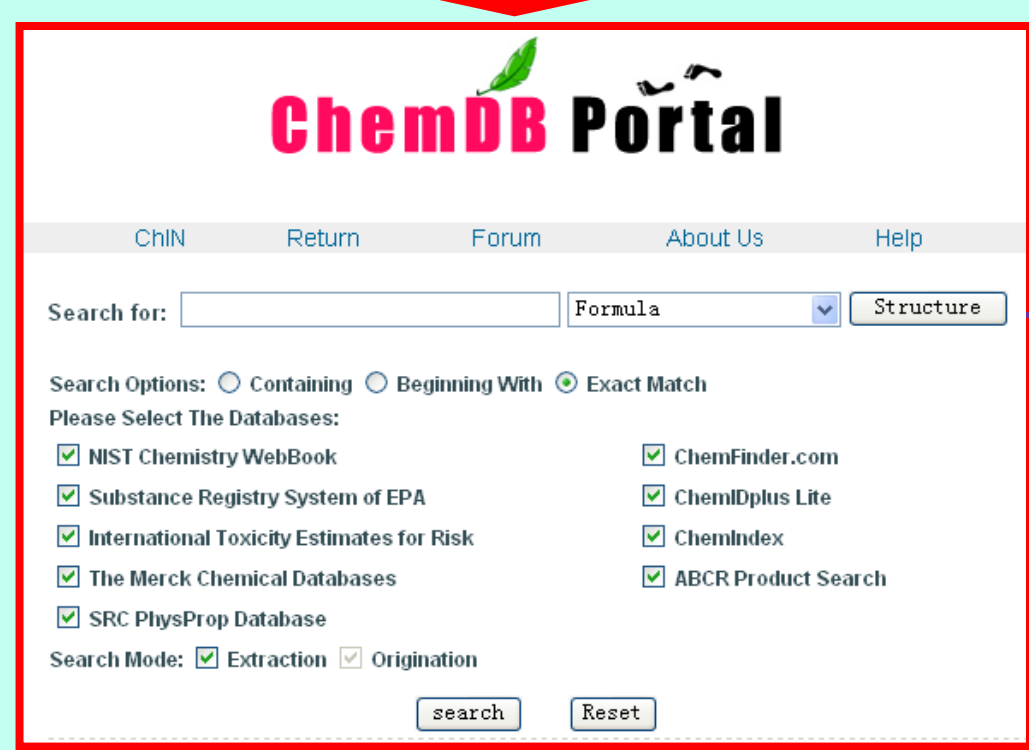
化学深层网数据挖掘与提取



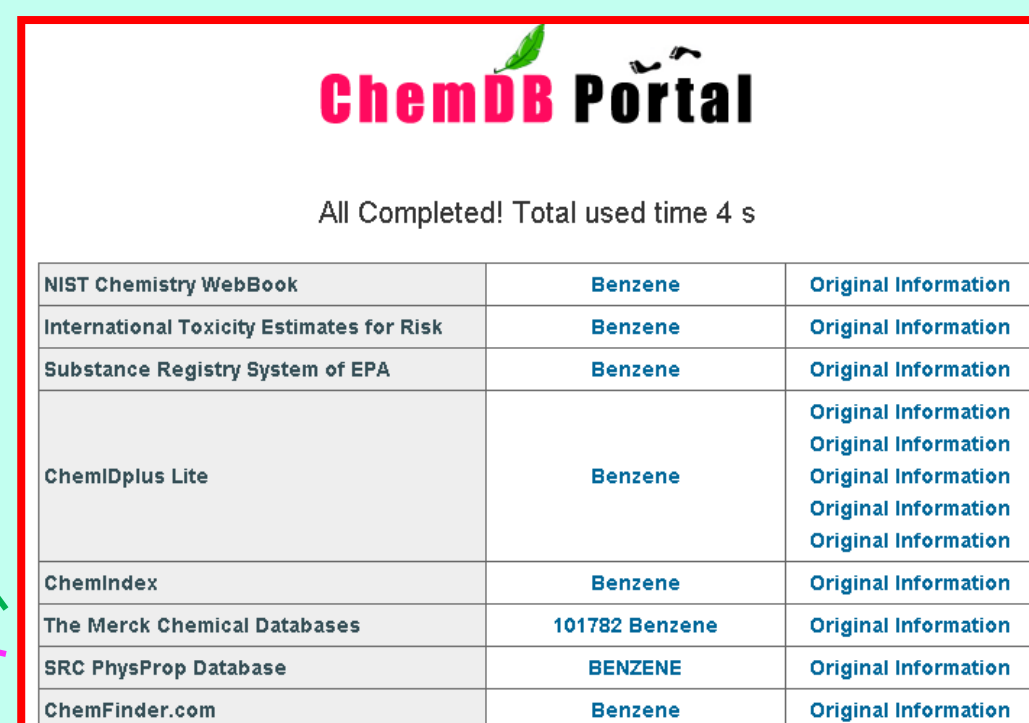
基于XML技术对分布在Internet上的化学化工数据库进行统一检索和数据提取, 已建立化学深层网检索引擎原型系统ChemDB Portal。



半自动数据提取模板生成工具



化学结构检索



多来源数据规范化输出