面向双一流建设评价体系的大数据平台研究

王全玉

北京理工大学计算机学院
School of Computer Science and Technology,BIT

内容提要

- 总体建设目标
- 主要特点
- 系统组成
- 数据来源
- 技术支撑
- 实现方法
- 部分成果

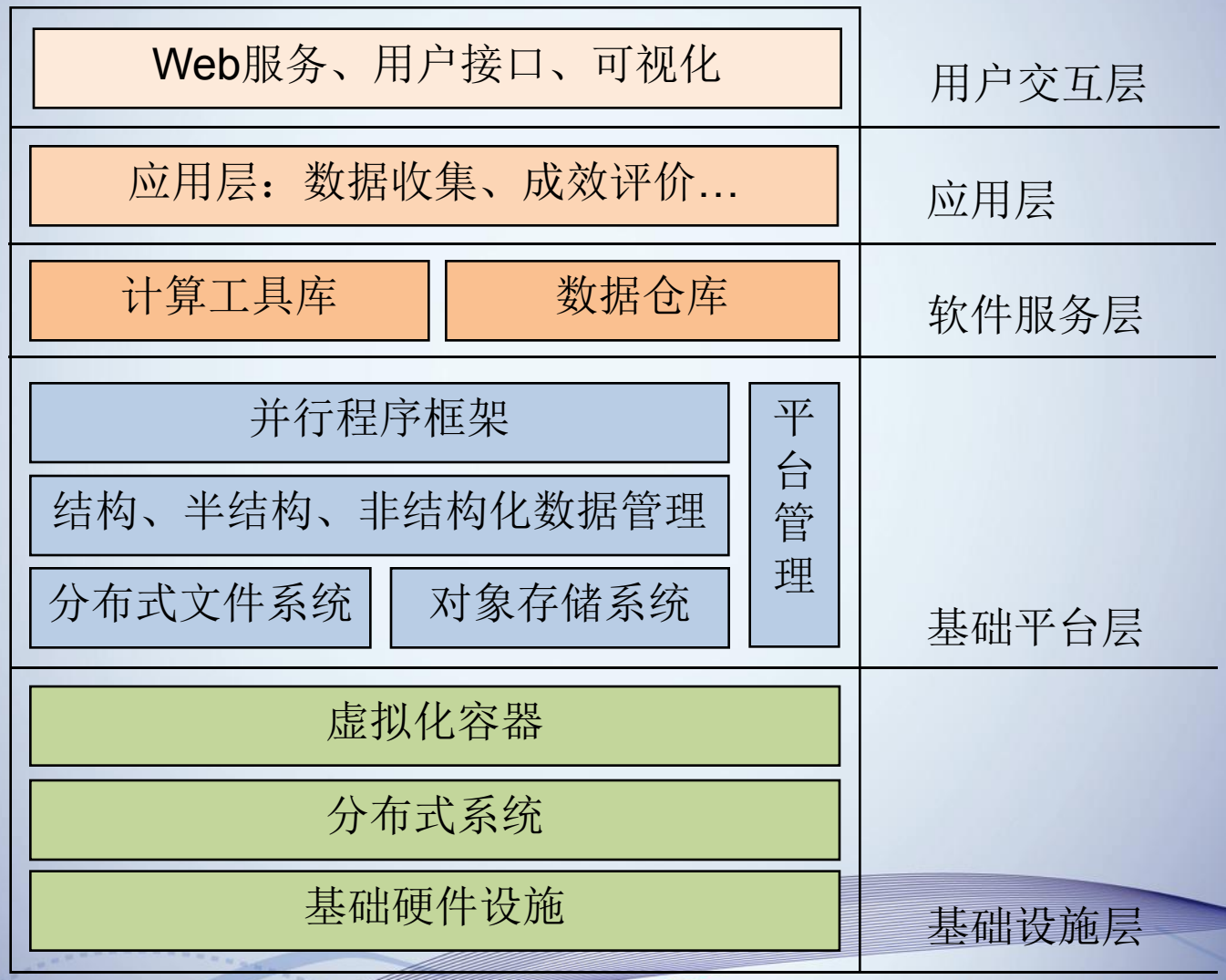
总体建设目标

- 目标：
  - 用数据说话，呈现”双一流”建设状态，为多元价值判断提供基础数据和计算平台
- 主要功能：
  - 状态呈现
  - 数据查询
  - 统计分析
  - 监测预警
  - 质量报告

主要特点

- 可扩展性强
- 数据处理能力强
  - 结构化数据
  - 半结构化数据
  - 非结构化数据
- 预警-一网打尽、一目了然
- 分析-定性定量地统计数据
- 管理服务-自动报告

系统结构图


Web服务、用户接口、可视化
用户交互层
应用层：数据收集、成效评价…
应用层
计算工具库
数据仓库
软件服务层
并行程序框架
结构、半结构、非结构化数据管理
分布式文件系统
对象存储系统
平台管理
基础平台层
虚拟化容器
分布式系统
基础硬件设施
基础设施层

学校　省/市教育主管部门　国家教育主管部门　评估机构/专家　社会公众

用户接口

双一流建设评价体系大数据平台

多维数据分析
与数据挖掘

主数据库
（数据仓库）

ETL(数据清洗与萃取)

全国高校教学基本状态数据库系统

满意度调查系统

公共信息导入系统

国际评估数据系统

各个高校　学生　用人单位　其他公共数据库　国际有关组织

北京理工大学计算机学院
School of Computer Science and Technology,BIT

组成部分

- 平台包括：
  - 可扩展的多个节点的平台
  - 多模数据库
  - 算法库
    - 通用算法库
    - 用户算法库
  - 可视化模块

组成部分

- 用户界面分为三类：
  - 在技术管理层面，偏重于建设。针对研究人员设置数据接口，可直接导入数据
  - 在信息使用层面，真正实现共建共享。该层用户可以建立和使用平台数据库、平台算法库，平台用户使用过的数据和算法进行累计和共享
  - 在数据发布层面，使用层面计算和处理结果公开后，一般用户可以通过图、表、等多种形式浏览结果，力争做到形式的多样化

主要监测内容

- 五个度
  - 达成度、贡献度、支撑度、影响度、引领度
- 监测内容
  - 五个建设任务
  - 五个改革任务
- 功能：
  - 监测项目：10个项目-建设+改革
  - 核心要素：24
  - 监测点：64

数据主要来源

- 动态性
  - 动态数据监测
  - 动态数据采集
  - 动态指标监测
  - 动态评价结果
  - 动态提醒预警

技术支撑

- 大数据、互联网+、云计算、人工智能
- WEB 挖掘技术
- 搜索引擎技术
- 自然语言处理技术
- 深度学习技术

数据采集

- 规范数据
  - 填报
  - 导入
  - 转换
  - 接入

数据采集

- 不规范数据
  - 网络信息雷达，定向采集和全网搜索结合
  - E（Extract）
  - C（Clean）
  - T（Transform）
  - L（Load）

实现方法

- 数据驱动
  - 数据是发动机，推动系统运行
- 基于组件的定义和定制
  - 计算模块组件化，通过定义和组装完成特定功能
- 监测嗅探
  - 基于自然语言理解和情感计算

实现方法

- 浓缩海量信息抵抗数据爆炸
  - 从教育质量的角度看，进行剔除和浓缩
- 强化数据挖掘实现信息增值
  - 通过对数据进行加工，对数据进行生产、分析和解读
- 跟踪关联数据提高趋势研判
  - 从注重静态收集向注重动态跟踪拓展
- 服务报警
  - 通过对已加工数据的定义、选取提供各种服务

实现方法

- 非规范数据辅助分析
  - 关于学校、教师、学生等方面的非结构化数据
    - 学校、教师、学生画像
    - 动态数据：教师或学生在竞技竞赛、学术活动等的活跃程度
    - 热点事件：如北大校长林建华讲话、清华大学胡鞍钢事件
  - 从个体、个例来说可能是不准确的，但如果数据量够大，宏观上大体是准确的

实现方法

- 算法库
  - 聚类算法
  - 分类算法
  - 推荐算法
  - 降维算法
  - 优化算法
  - …

实现方法

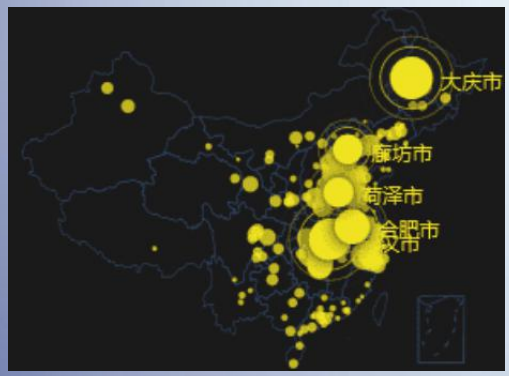
数据挖掘	机器学习	语言处理
• K-means算法	• 决策树学习	• 自然语言理解
• 支持向量机	• 最大期望算法	• 信息抽取
• 朴素贝叶斯	• 深度学习	• 情感计算
• …	• …	• …

数据结果呈现

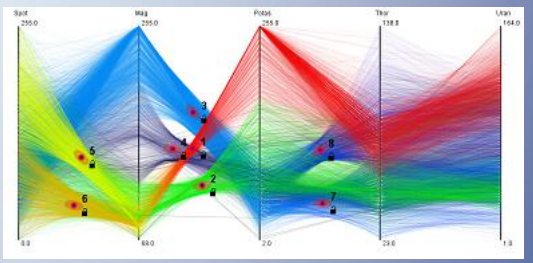
经典图表

交互仪表

热力地图



多维可视



大数据在教育领域的应用

大数据在教育领域词云图