

大数据环境下的地质资料 智能服务技术

陕西省地质资料档案馆
高洋

引用格式请参考：高洋.大数据环境下的地质资料智能服务技术(V1).
陕西省地质资料档案馆[创建机构], 2020.全国地质资料馆[传播机构],
2020-5-22. doi:10.37088/ngac.ppt10

目录

01 大数据环境的形成与特点

02 数据处理与服务技术

03 地质资料智能服务思考



1. 大数据

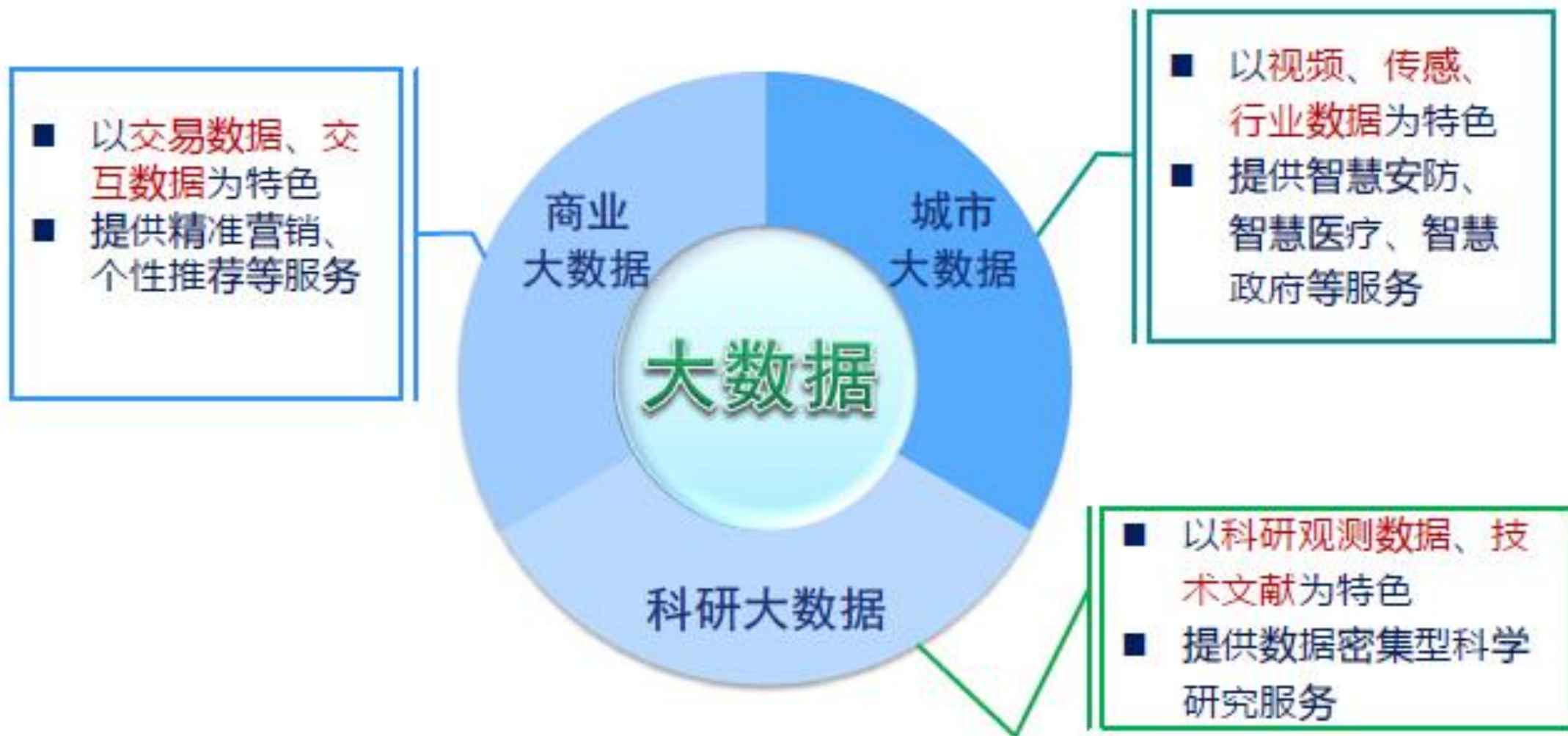
大数据

大数据是指在可容忍的时间内无法用现有的信息技术和硬件工具对其进行传输、存储、计算与应用等的数据集。

- 数据量大 (Volume)
- 类型繁多 (Variety)
- 价值密度低 (Value)
- 速度快时效高 (Velocity)



1.1.大数据应用领域



1.2.大数据产业政策

2015年8月，国务院印发了《促进大数据发展行动纲要》；

2016年12月，工信部编制印发了《大数据产业发展规划（2016—2020年）》；

2017年7月，国务院印发了《新一代人工智能发展规划》，系统部署大数据和人工智能发展工作。

大数据产业发展规划 (2016-2020年)

新一代人工智能
规划布局图



1.3.地质资料大数据

地质资料是地质工作的主要成果，是在长期的基础地质调查、环境与地质灾害监测、资源调查评价、地质科技支撑、地质数据更新与应用、海洋地质等调查与研究的过程中辛勤劳动的结晶。

与结构化数据数量相比，非结构化数据在地质资料中所占比例要比结构化数据大得多，所涉及的内容更丰富，潜在价值也是其他数据无法比拟的。



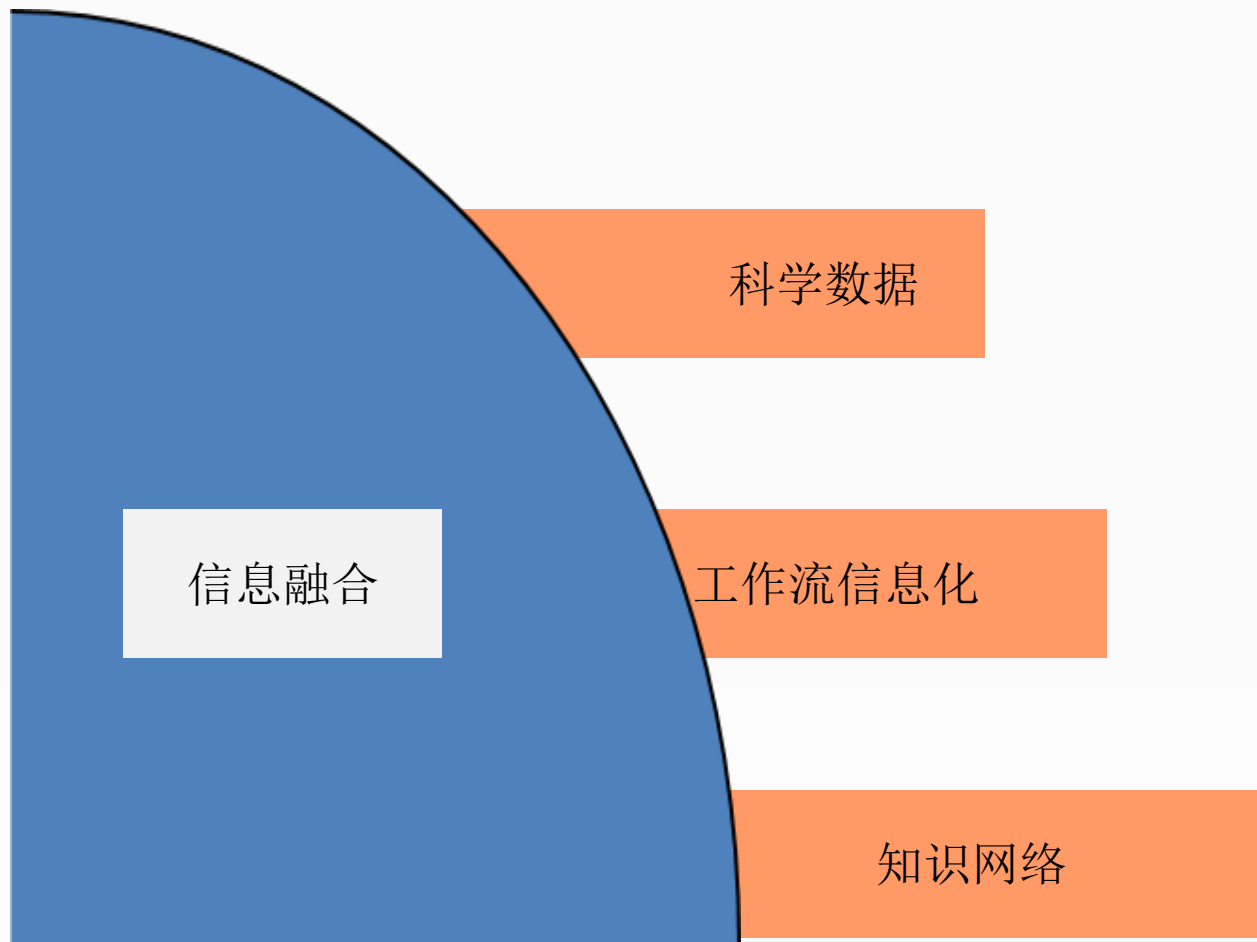
- **巨量化、共生性、碎片化：**数据总量庞大，馆藏机构的地质资料存储空间在TB级别；数据零散分布，与结构化数据（地质数据库）共存共生；文件“小而碎”，一部报告可由上百个文件乃至上千个文件构成：正文类、附表类、附件类、附图类、多媒体类、审批类、其他类。
- **多样性：**格式多样性，Word、PDF、Excel、PPT、图形、图片、视频等；内容多样性，数据之间关系复杂，不容易表达。
- **保存方式简单：**采用科技档案通用的案卷方式保存，对资料内容揭示有一定限制，无法详细地显示每档资料包含的信息，存在大量隐藏信息不易获取等问题。
- **价值储量高、信息分散：**内容组织灵活，多种价值信息交织在一起，信息总量大，但是某一种类的信息比较分散。
- **历史积累：**数据更新频率不高，历史积累的数据内容代表了一块区域或是机构数年甚至几十年积累的知识 and 智慧。
- **时空特征明显：**数据本身与空间位置有紧密的联系，如：矿产资源的分布规律；数据也具有时间属性，如：不同地质时代的岩石、地层、矿床具有不同的分布特征和规律。



2.大数据发展趋势

当前的自然资源知识与信息服务的环境正在发生变化：

- 1.电信网、计算机网和有线电视网三大**网络融合**成为一种趋势；
- 2.整个社会的教育和知识水平显著提高，组织、机构和个人都掌握一定数量的知识资源，形成了一种**泛在**的知识环境；
3. **数据科研模式**（第四范式），对海量信息产生了新的需求；
- 4.互联网及相关数据库服务逐渐向**知识化、智能化转型**，满足新的知识需求。



目录

01 大数据环境的形成与特点

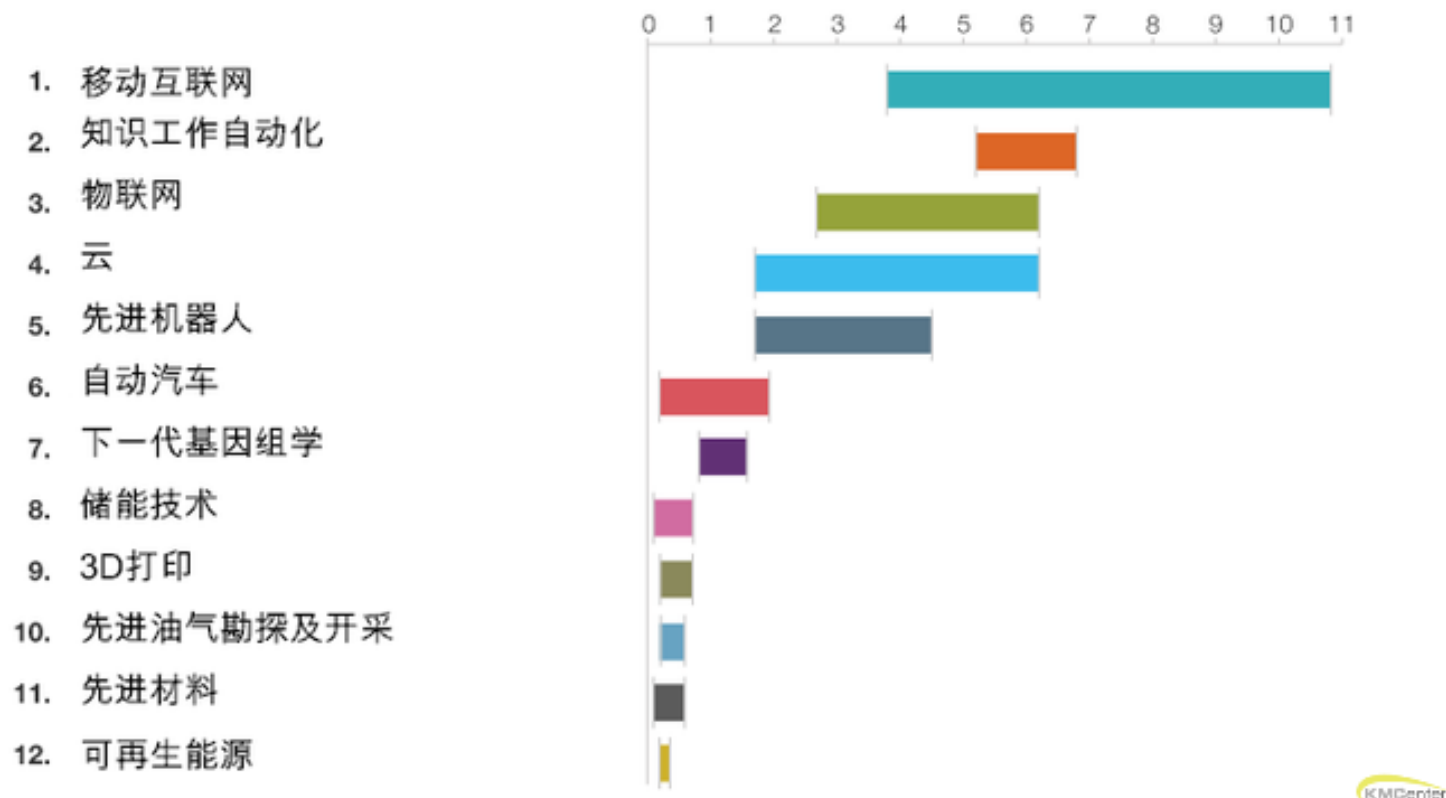
02 数据处理与服务技术

03 地质资料 智能服务思考

1.数据处理技术发展前景

颠覆技术图示

至2025年的预估潜在经济影响上下限（万亿美元，年度）



麦肯锡展望2025：决定未来经济的12大颠覆技术

2. 数据处理技术应用



#2 Automation of knowledge work

Intelligent software systems that can perform knowledge-work tasks

Potential economic impact in 2025 across sized applications of
\$5.2 trillion–\$6.7 trillion

Additional labor productivity could equal the output of **110 million–140 million** full-time workers

Component technologies

- Artificial intelligence, machine learning
- Natural user interfaces
- Big-data technologies

Key applications

- Smart learning in education
- Diagnostics and drug discovery in health care
- Discovery, contracts/patents in legal sector
- Investments and accounting in finance sector

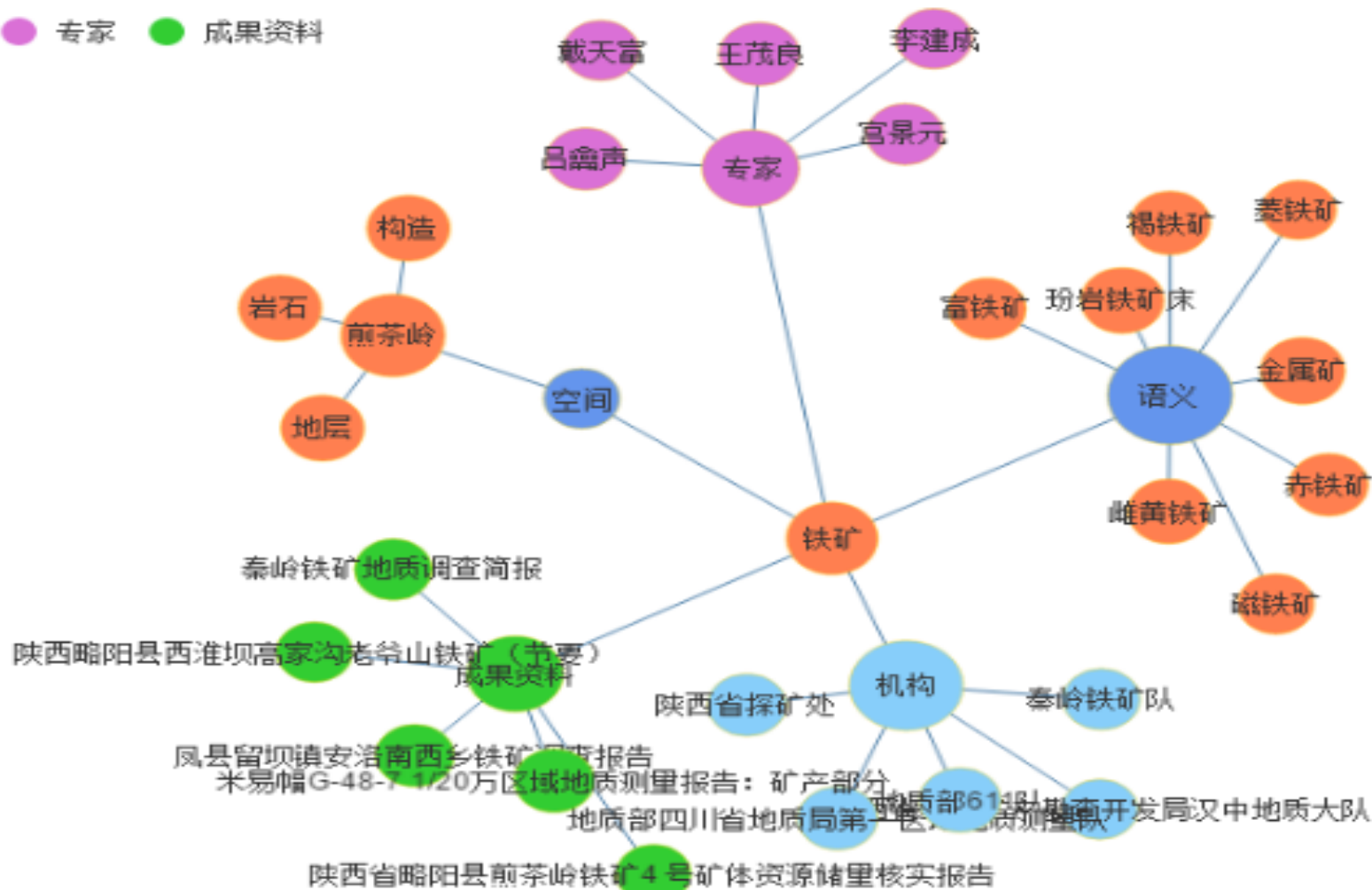
3.知识工作自动化

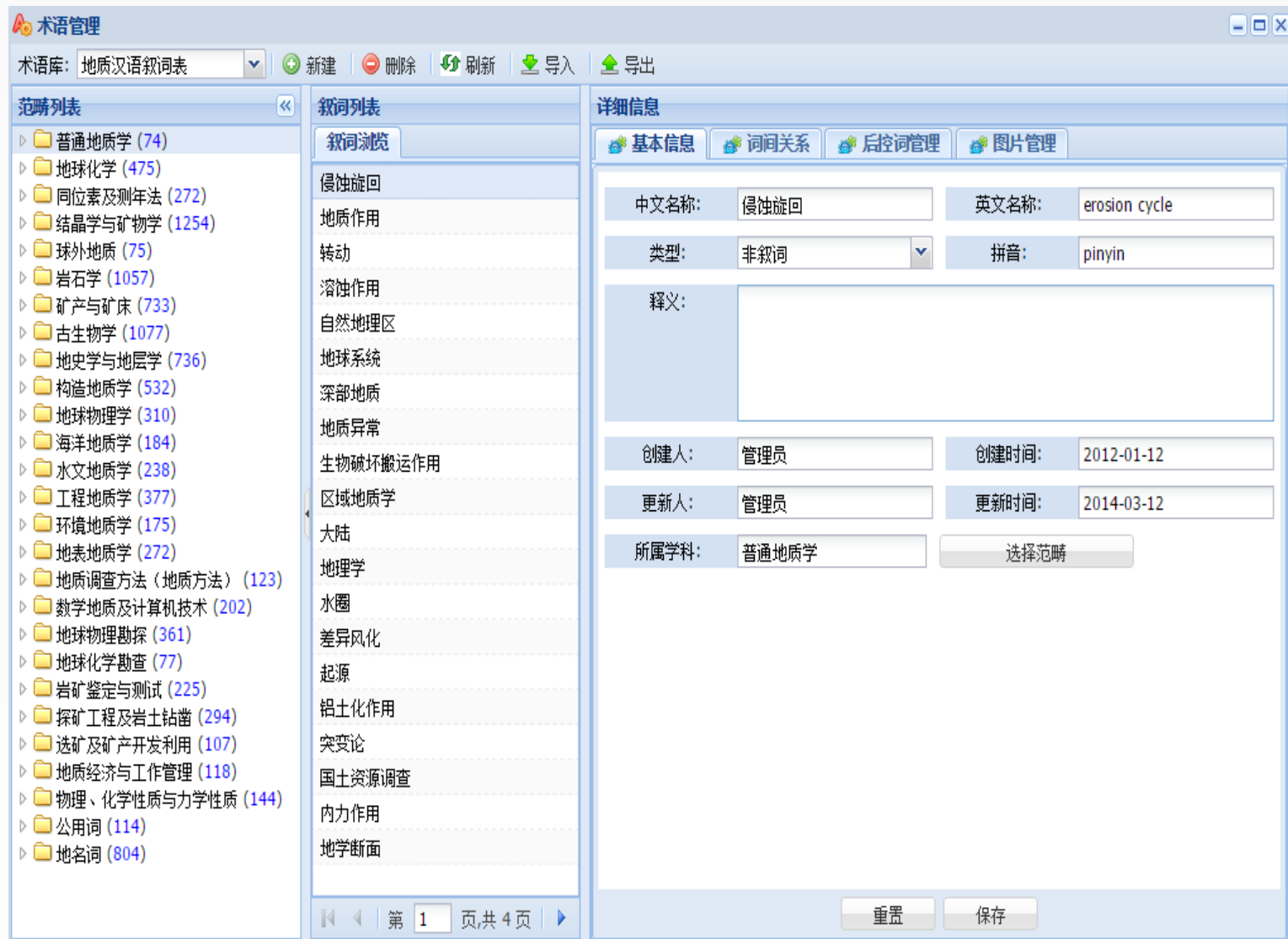


- 人机交互：自然语言处理——专业领域词系统
- 大数据技术：文本结构化与数据关联
- 人工智能：从信息服务到智能服务

3.1.自然语言处理——专业领域词系统

● 主题 ● 机构 ● 专家 ● 成果资料





词系统术语的三个特点:

- ◆ 一是概念是唯一的
- ◆ 二是术语间语义关系
- ◆ 三是记录了术语的历史演化



3.2.大数据-文本结构化处理-知识元

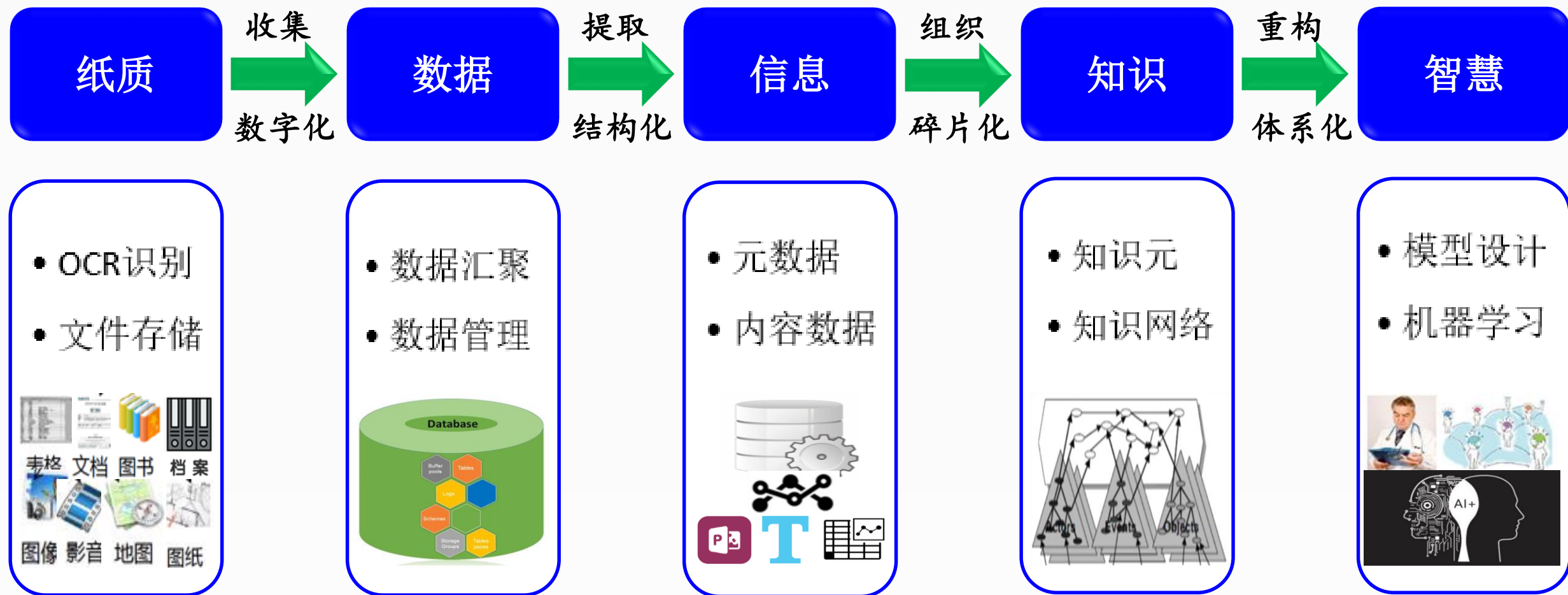
非结构化文本在结构化的过程中，就是一个知识体系重新架构的过程。

文本结构化处理的两大技术领域：

- ◆ 一是知识内容的结构化表示方法：知识元——知识表示；
- ◆ 二是知识结构化技术：文本识别——细粒度存储——结构化关联。

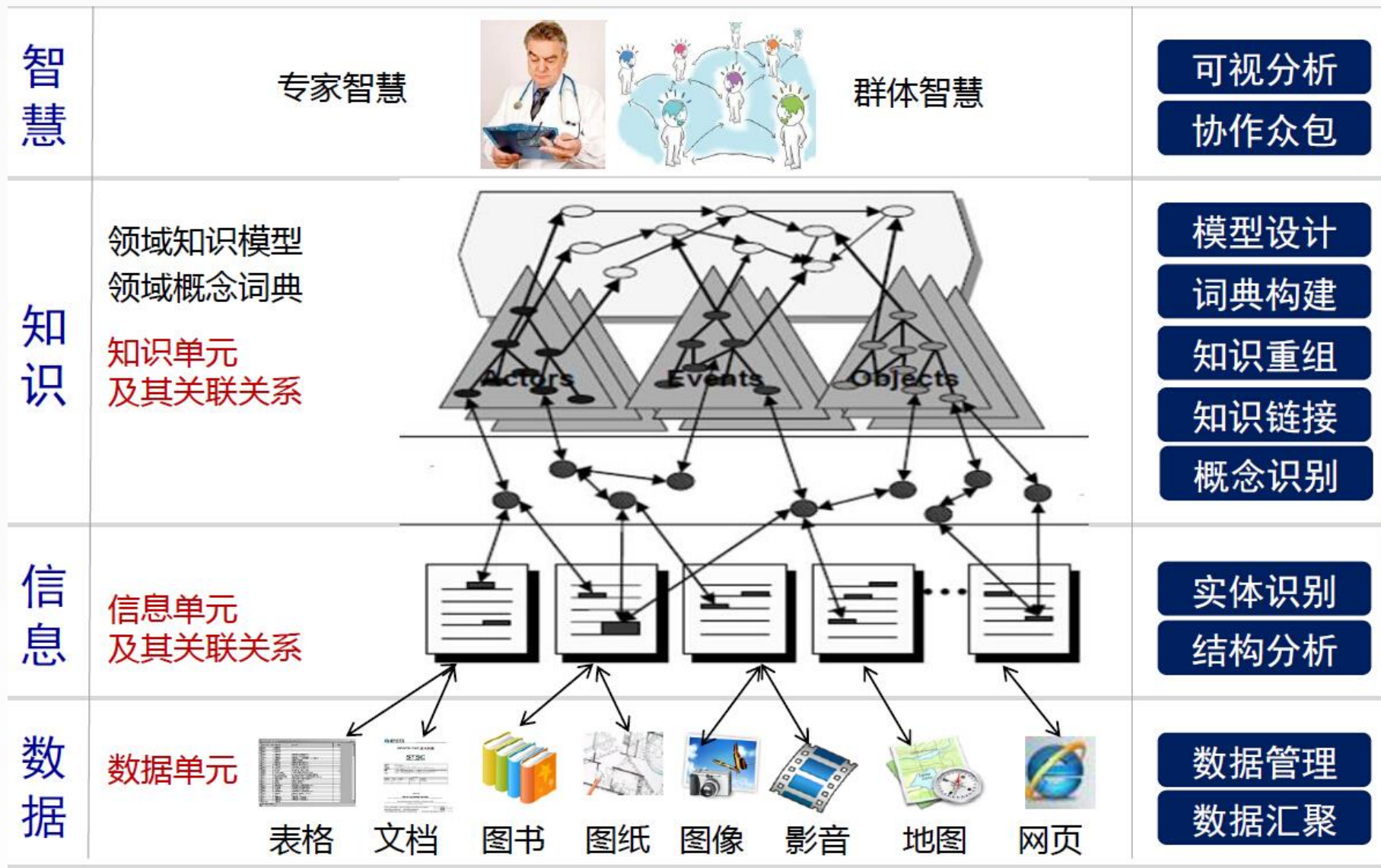
知识元理论：

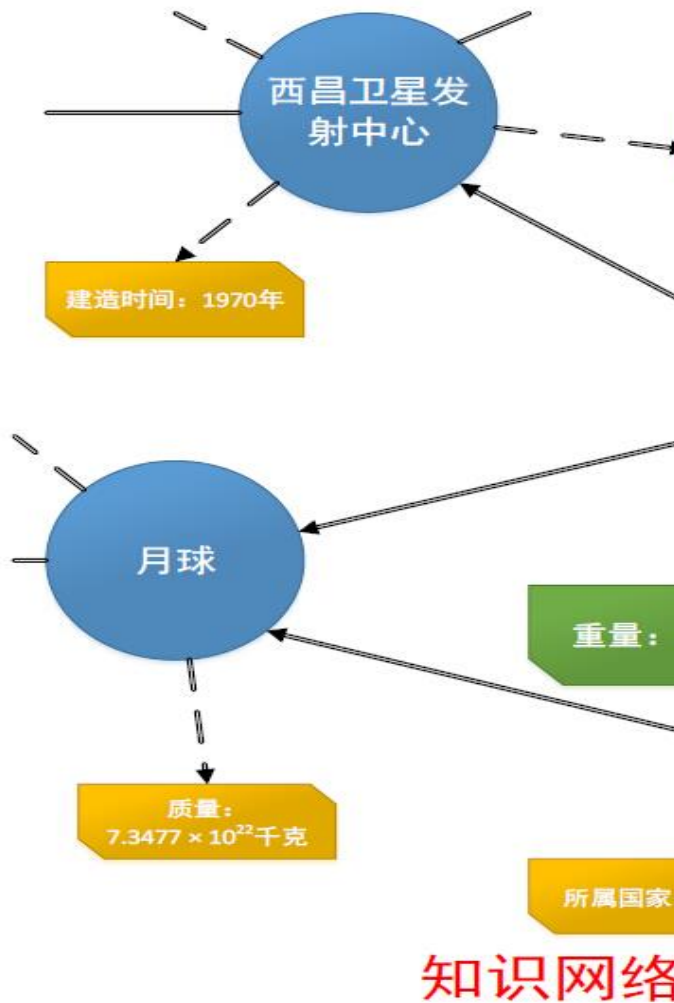
- ◆ 一是知识内容视角，提出了文献单元、信息单元和知识单元的概念；
- ◆ 二是生物基因视角，将知识元看作是决定知识体系生长演化的最小控制单元；
- ◆ 三是系统工程视角，将知识元看作是网络节点。



数据结构化处理技术过程

3.3. 智能服务





陕西省地质资料集群化智能服务系统

当前用户：游客 登录 注册

首页 地质资料 钻孔 矿业权 储量 科技文献 术语资源 地方志

所有资源 ▾ 搜索

铁矿 煤矿 开发利用 核查报告 监测说明

第一章 概况

一、目的任务

为了加强矿山储量管理，摸清资源家底，建立矿山储量动态监督管理制度，促进资源合理利用与保护，并为采矿权价款评估和资源储量登记、统计提供依据，根据陕西省人民政府办公厅陕政办发[2004]121号文和陕国土资发[2006]25号文要求，由镇安县国土资源局委托陕西省地质矿产综合地质大队对月西硫铁矿的资源储量进行检测，主要任务如下：

- 1、分割计算矿山（采矿证范围内，下同）占用国家出资勘查形成矿产地的资源储量（累计动用量、剩余量及占用查明资源量）；
- 2、估算矿山建设后开采动用的资源储量（含开采量、损失量、勘查或计算增量）；
- 3、核实矿山保有和累计查明资源储量；
- 4、调查、评价矿山资源开发利用中存在的问题，并提出相应意见或建议。

二、位置、交通及自然地理

月西硫铁矿位于镇安县城170°方位33km处，行政区划属陕西省商州市镇安县青钢关镇管辖。矿区中心地理坐标为东经：109°11'40"，北纬33°13'50"。矿山位于青钢关镇东侧3km，有村、镇水泥公路相连，西康铁路、西康在建的高速公路从青钢关通过，交通方便（见插图1-1）。自然地理属秦岭山脉南坡汉江流域乾佑河中游，地表为沟谷地貌，切割强烈，地形陡峻，最大海拔高程1935.1米，最小海拔高程423.8m，地势总体东高西低。区内属大陆性温带半湿润气候，年平均气温13.5℃，极端最高气温42℃，极端最低气温-8℃。年平均降水量845.0mm。

第一章 概况

一、目的任务

二、位置、交通及自然地理

三、以往地质工作及资源储量评审

四、矿权设置

第二章 本次检测工作及质量评述

一、检测工作阶段

二、采空区测量

三、矿山生产探矿工作

四、检测工作质量评述

第三章 地质特征

一、地层、构造与岩浆岩

二、矿体地质特征

三、矿石质量

四、开采技术条件

第四章 矿山开采及资源利用情况

一、矿山建设和开采情况

二、矿山开采对地质环境的影响

三、探采对比

四、越界开采情况

第五章 矿山开采技术经济分析

一、市场分析

二、矿山建设内、外部环境评述

三、矿山生产现状

四、技术经济分析

目录

01 大数据环境的形成与特点

02 数据处理与服务技术

03 地质资料智能服务思考



1.深度典型需求

1)专业领域

- ◆ 重大科学、工程问题的多来源信息集成;
- ◆ 文本的结构化呈现;
- ◆ 专题术语知识清理;
- ◆ 文本细粒度内容重组。

2)信息组织部门

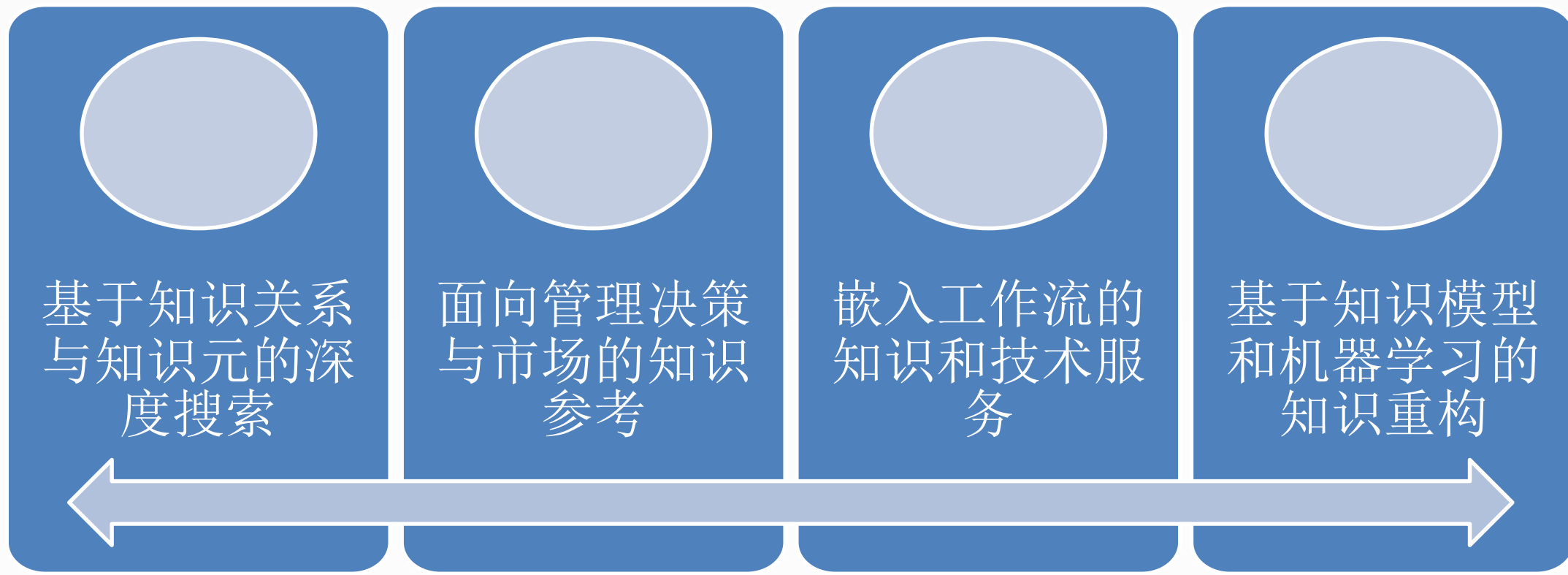
- ◆ 信息的结构服务;
- ◆ 数据标引服务;
- ◆ 知识化内容服务。

3)管理与决策部门

- ◆ 决策知识参考;
- ◆ 决策数据支持（嵌入过程的内容服务）。

2.智能服务展望

我国已经建立了从国家到地方的地质资料馆藏体系。地质资料按所载信息是否涉及国家秘密分为非涉密地质资料和涉密地质资料，馆藏机构根据用户涉密类别向其提供相应地质资料。用户可以通过网络、到馆申请查阅、复制所需的地质资料。



Thank You