



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

基于医学科技查新词频可视化分析的

图书馆学科服务创新研究

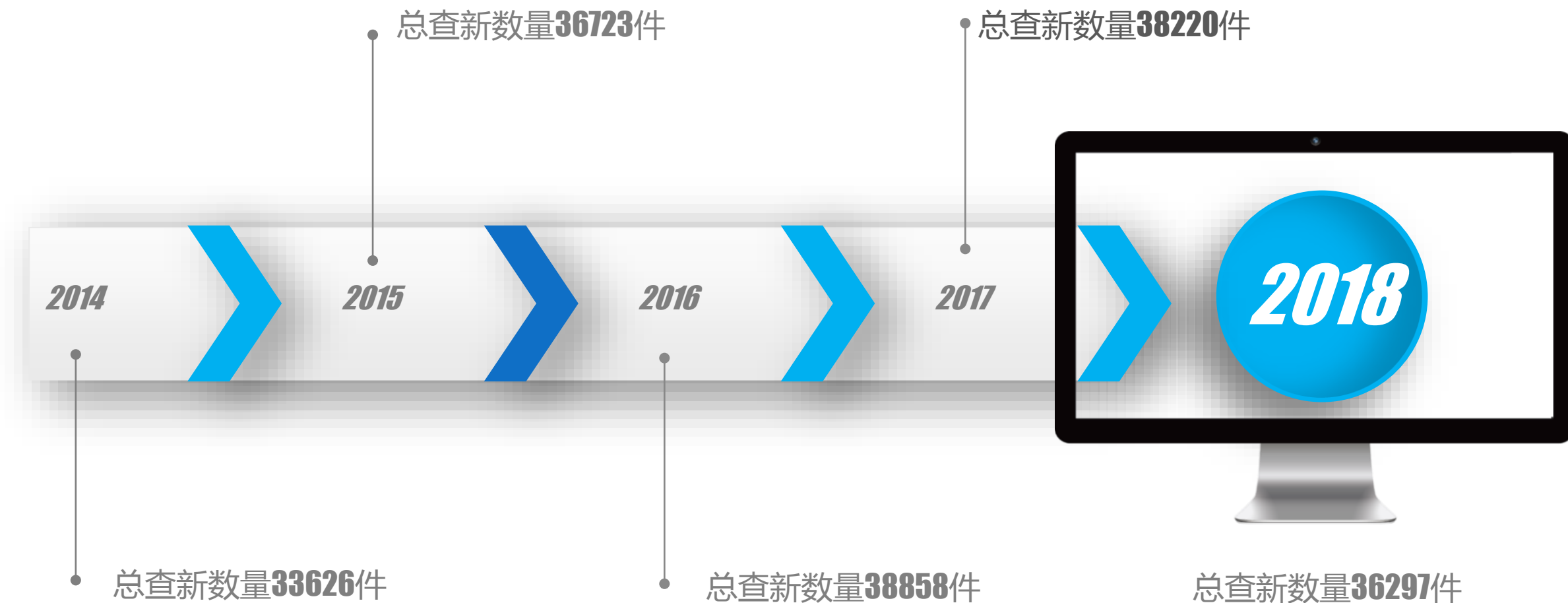
 2018年CALIS医学文献信息中心重点科研基金项目研究成果汇报

 西安交通大学图书馆 杨超

# 选题背景



西安交通大学  
XI'AN JIAOTONG UNIVERSITY



近五年全国102所教育部查新站平均查新数量为36745件



- ✓ 如何发挥查新报告的潜在价值？
- ✓ 如何结合查新数据挖掘学科领域热点？
- ✓ 可否为科研学者申报项目提供参考信息？
- ✓ 研究方法能否应用于其它学科？

# 供需分析



西安交通大学  
XI'AN JIAOTONG UNIVERSITY



## 一锤定音

查新站被动接收查新请求，完成查新报告



## 角色转换

图书馆**主动**嵌入学科服务，为学者提供学科情报分析

## 基础服务

项目立项  
课题结题  
申报奖项  
申请基金  
不良反应

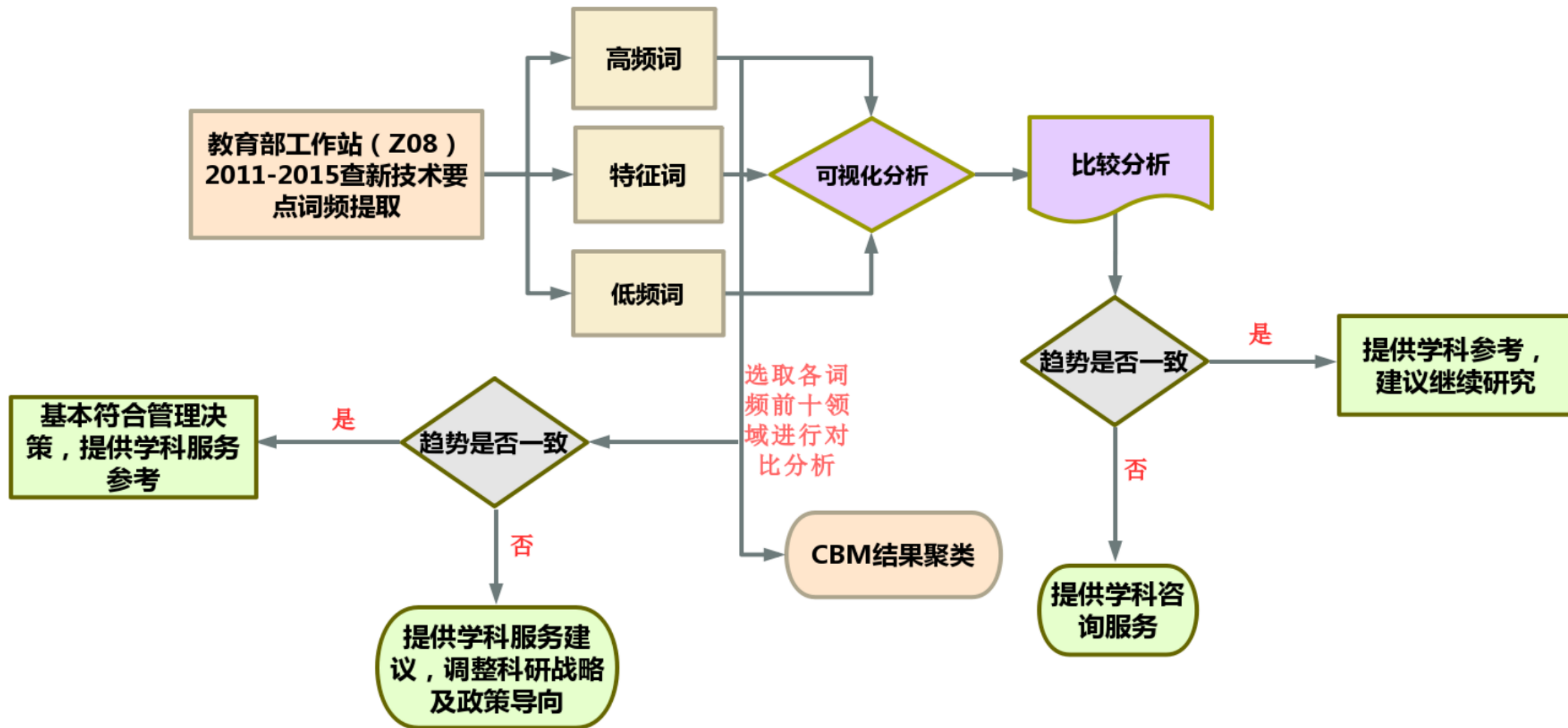


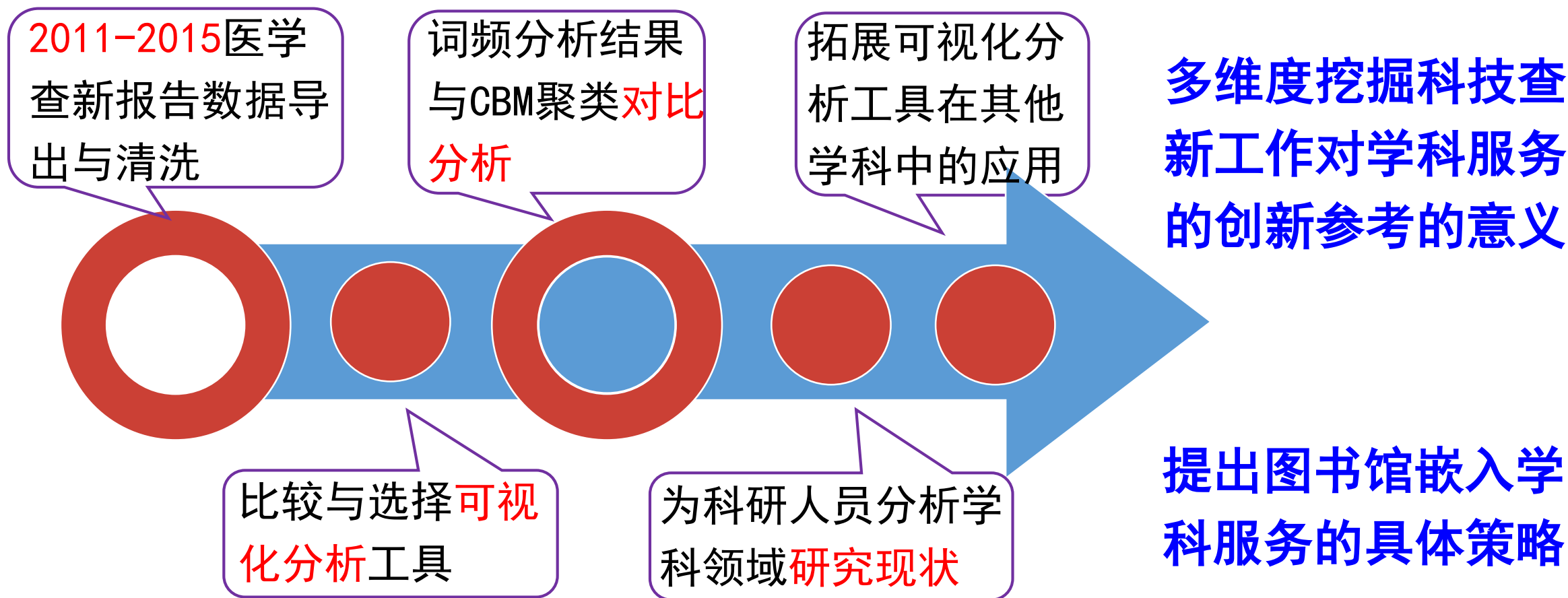
## 学科服务

挖掘潜在需求  
提供前沿热点  
鉴别当前盲点  
**指点项目选题**  
**分析学科动态**



# 方案设计





# 研究重点及创新点



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

可视化工具的选择及优劣势分析



拓展可视化分析工具  
在其他学科中的应用



结合词频分析结果  
提出专业性建议

## 在线数据平台

知网指数分析

万方知识脉络分析



知网图片分析

维普科学指标分析

知网可视化分析



## 在线数据平台前期研究

### 知识脉络分析

——基于万方的在线分析工具

杨 超,雷润玲,张幸芝

(西安交通大学 图书馆 陕西 西安 710061)

**摘 要** 从知识脉络分析平台的基本功能、进入方式与分析界面、知识脉络检索分析、相关词比较分析等四方面,详细介绍了基于万方数据知识服务平台上的在线分析工具——知识脉络分析,揭示了多维度的可视化分析,对学术选题、学术研究等方面均有较大的参考意义。

**关键词** 知识脉络分析 万方 在线分析工具

**中图分类号** G252.7 G254.97

### 指数分析

——基于 CNKI 的在线分析工具

雷润玲,杨 超,张幸芝

(西安交通大学 图书馆 陕西 西安 710061)

**摘 要** 介绍了基于 CNKI 的指数分析工具,专门针对 CNKI 在线分析工具进行了图解,以帮助读者了解、掌握、运用工具,解决文献利用过程中,尤其是文献分析过程中遇到的疑问或困难。

**关键词** 指数分析 CNKI 在线分析工具

**中图分类号** G252.7 G254.92

### 图片分析——基于 CNKI 的在线分析工具

雷润玲,李 凌,杨 超,张幸芝

(西安交通大学 图书馆 陕西 西安 710061)

**摘 要** CNKI 图片在线分析工具主要有检索、比较、分析和以图搜图 4 项基本功能,文章主要从图片颜色、学科、发表年度、关键词、图片类别 5 个角度对图片检索结果进行了分析,以帮助读者尽快了解和认识所关注的图片,了解和掌握所关注图片的相关情况。

**关键词** 图片分析 CNKI 在线分析工具

**中图分类号** G254.97

### 科学指标分析

——维普资源整合平台的在线分析

李 凌,雷润玲,杨 超,张幸芝

(西安交通大学 图书馆 陕西 西安 710061)

**摘 要** 介绍了维普资源整合平台中科学指标分析数据库的概况,通过图解及实例演示,帮助科研工作人员了解、掌握、运用其分析工具,在科研过程中准确把握科研动态、研究热点和前沿,客观衡量国内科学研究绩效,跟踪国内科学发展趋势。

**关键词** 科学指标分析 维普数据分析 在线分析

**中图分类号** G254.97

## 文本挖掘

*CiteSpace*

UCINET

Thomson Data  
Analyzer<sup>®</sup> (TDA)

*ROST-CM*

# 主题词知识脉络分析方法



万方分析

主题分析

学者分析

机构分析

学科分析

期刊分析

地区分析

“探究主题领域知识脉络变化

用数据支撑主题研究

- 把握国内外主题研究走势
- 探究主题渗透的学科领域
- 发现主题相关的专家学者
- 了解主题相关的基金资助
- .....

排序	热词	出现频次
1	高校图书馆/图书馆	111
2	查新质量	63
3	检索策略	44
4	Sci-tech novelty retrieval/科技查新/科技创新	36
5	文献检索/信息检索/情报检索/光盘检索/检索/光盘	32
6	统计分析	32
7	中医药/医药卫生	20
8	信息服务	10
9	信息咨询/咨询中心	9
10	文献资源	9
11	数据库管理/数据库	8
12	江西省	7
13	查新报告	6
14	问题	5
15	intenet	4
16	Citespace/知识图谱分析/可视化分析	3
17	科研成果/成果鉴定	2
18	保障环节	2

# 主题词—高低频词提取方法



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## ROSTCM

### 词频分析的步骤

- (1) 多文本文件合并
- (2) 分词
- (3) 词频分析
- (4) 去除无意义词汇



# 主题词—特征词提取方法



## ROSTCM

特征值（TFIDF）是指某个词在一篇文献中出现的频率高，并且在其他文献中很少出现，则被认为此词具有很好的类别区分能力。

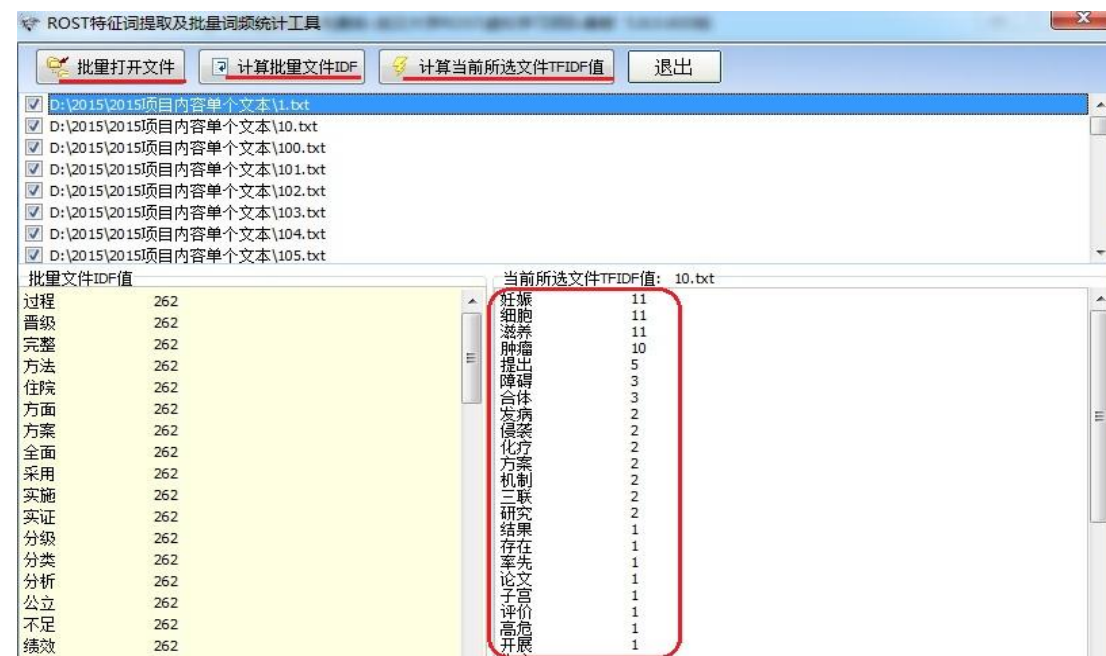
TFIDF 的计算公式为：

$$TFIDF_{i,j} = TF_{i,j} \times IDF_i$$

其中：

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

$$IDF_i = 1 + \lg \frac{|D|}{|\{j : t_i \in d_j\}|}$$



TF表示词频，IDF表示逆向文件频率，通过计算特征值由高到低排列所对应的特征词体现了学科领域特色水平。

- (1) 根据文本内容分成若干个独立文本文件；
- (2) 逐个计算文件的TF-IDF值。

# 主题词对比结果



## ROSTCM

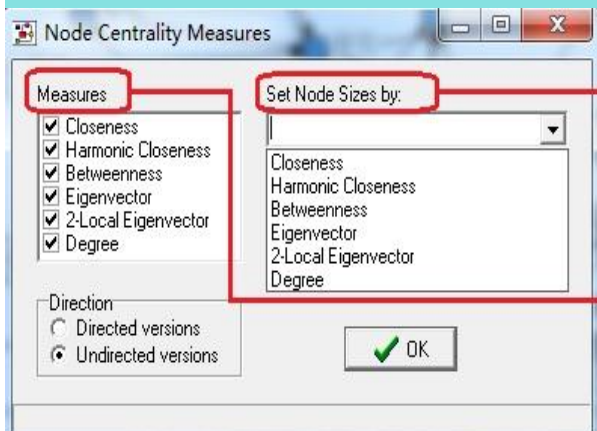
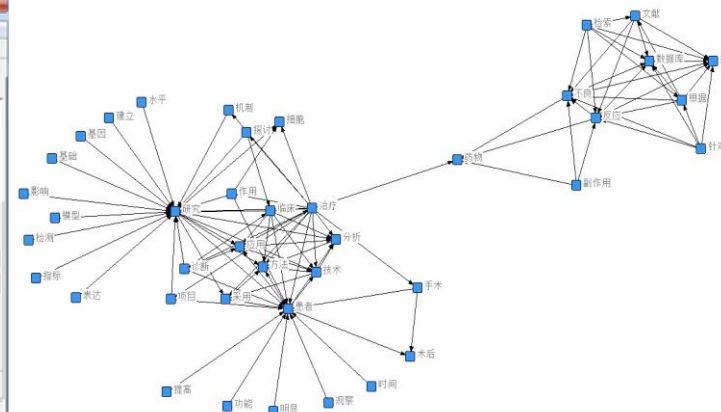
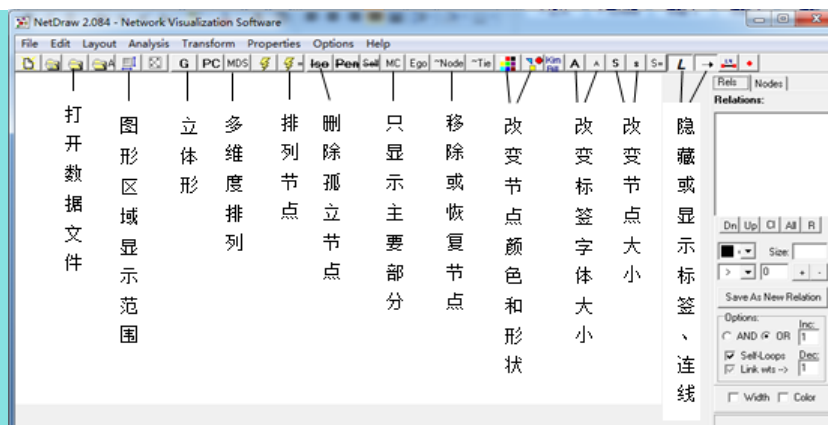
词语	词频	特征词	TF-IDF值	词语	词频	特征词	TF-IDF值
研究	475	损伤	51	方法	165	骨髓瘤	22
患者	462	移植	49	分析	162	子宫	22
治疗	410	抗体	40	肺癌	150	克山病	22
手术	301	检测	28	作用	148	辅助性	22
肿瘤	269	食品	26	不良	143	细胞	22
细胞	263	药品	24	损伤	142	突变	22
临床	246	穿刺	24	乳腺癌	139	装备	21
胃癌	211	痛风	23	术后	133	胸骨	21
反应	196	乳头	23	诊断	132	淋巴结	21
应用	192	四化	22	采用	129	转移	21

# 主题词社会网络分析方法



## NetDraw

借助ROSTCM可将文本构建成为NetDraw可识别的语义网络文件格式。直接选择“启动NetDraw”，可呈现出可视化网络语义的初始图谱。



设置节点大小取决于哪个要素

选择需要计算节点的哪些要素

要得到一个特点鲜明、便于分析的图谱，需要通过分析菜单里的中心性测量，即“Analysis” | “Centrality Measures”，为每个节点添加关键要素。

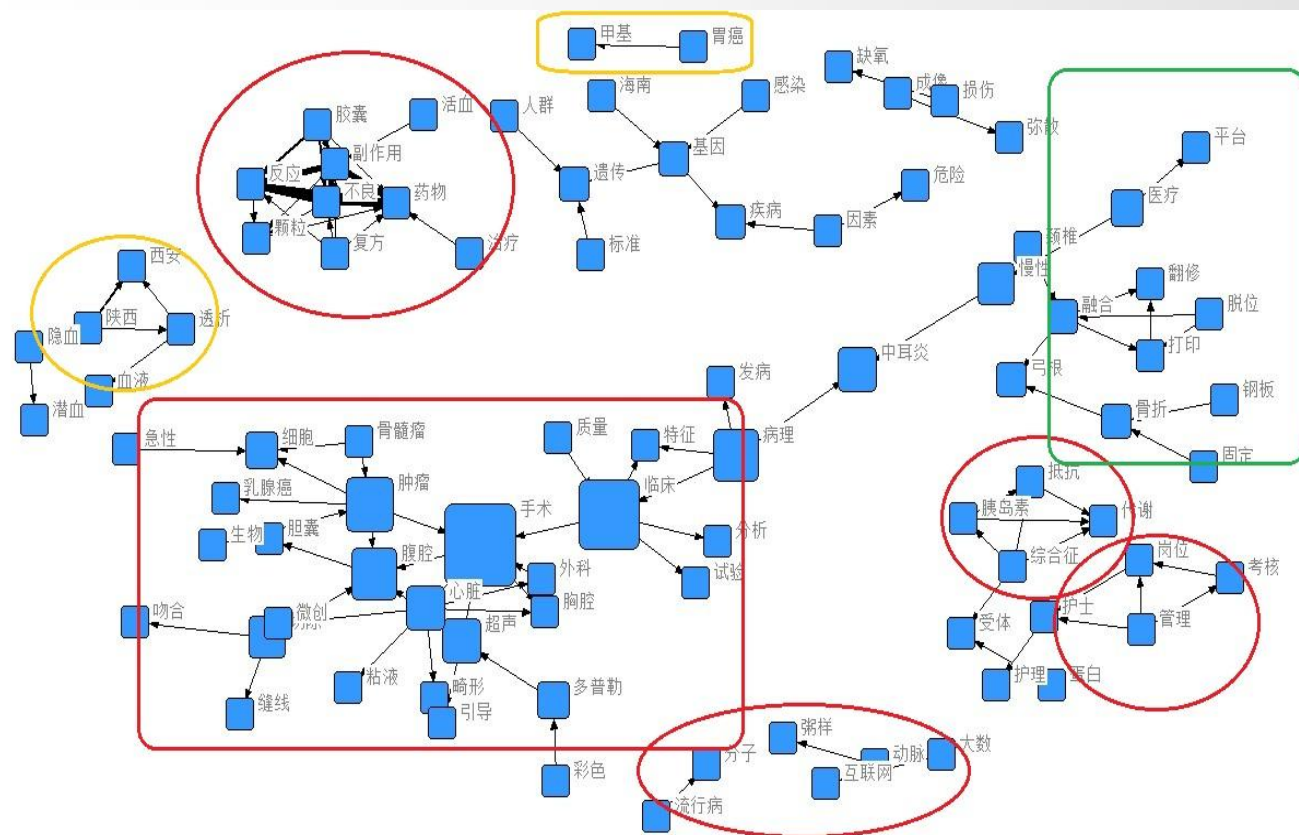
# 主题词可视化分析方法



## ROSTCM+NetDraw

查新关键词热点包括：

肿瘤  
外科手术  
药物副作用  
心脑血管  
骨折相关  
胃癌和甲基相关  
护士岗位考核管理等  
研究内容有些相互关联，有些独立存在，便于研究者快速梳理查新项目的内容脉络



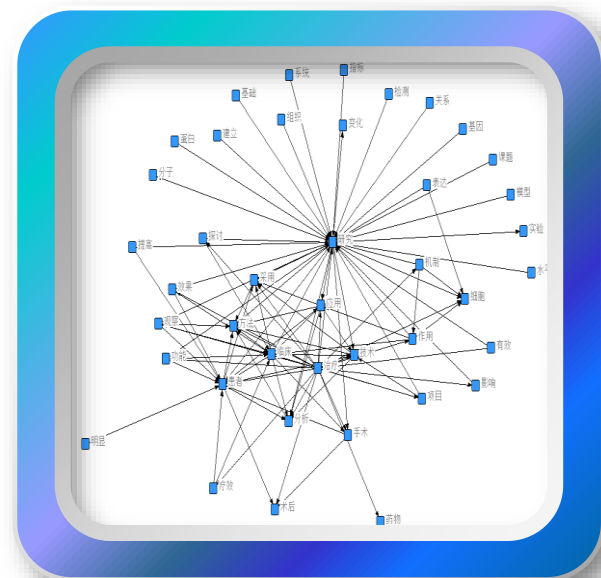


# 主题词可视化分析结果

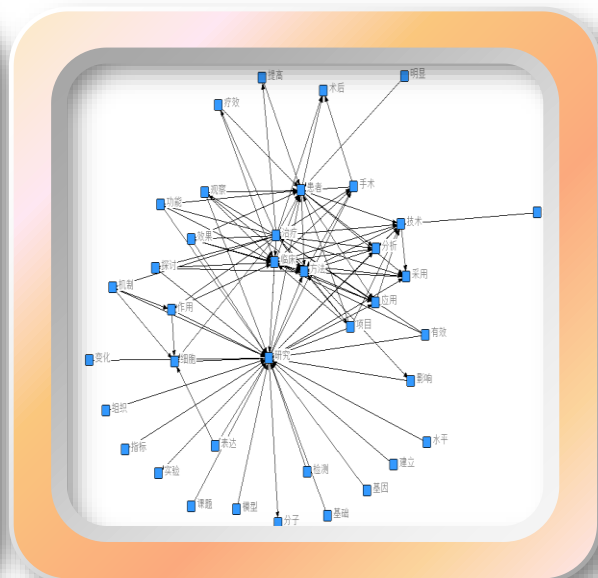


西安交通大学  
XI'AN JIAOTONG UNIVERSITY

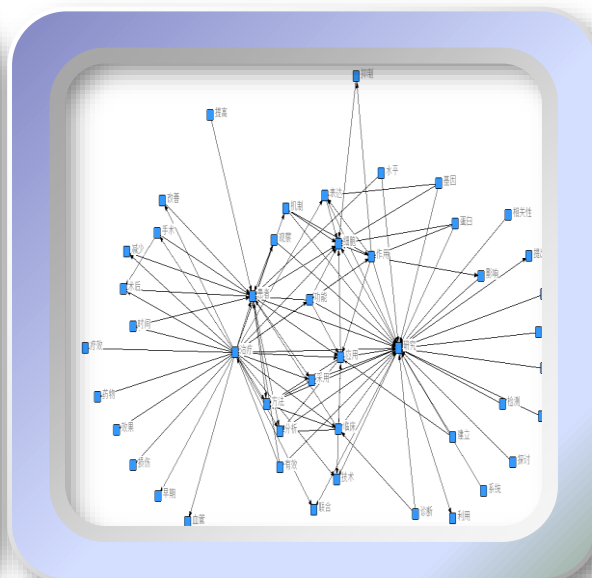
## ROSTCM+NetDraw



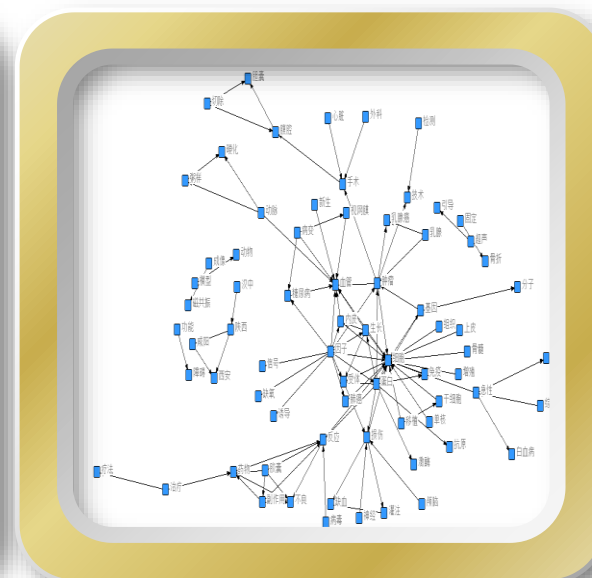
2011-2015年全  
部数据共现网络



2011-2015年项  
目内容共现网络



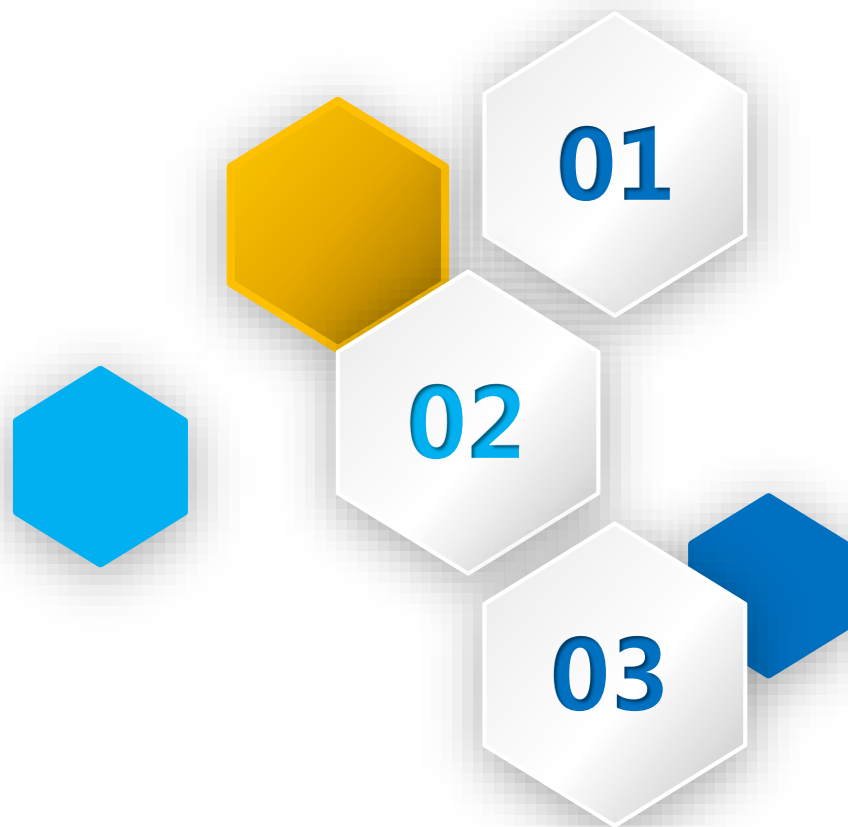
2011-2015年创  
新点共现网络



2011-2015年关  
键词共现网络

# CBM主题聚类方法





01 通过词频分析为学科领域提供政策导向参考

02 通过嵌入式学科化服务为科研人员提供信息参考

03 通过建立科技查新合作联盟拓展学科服务范围

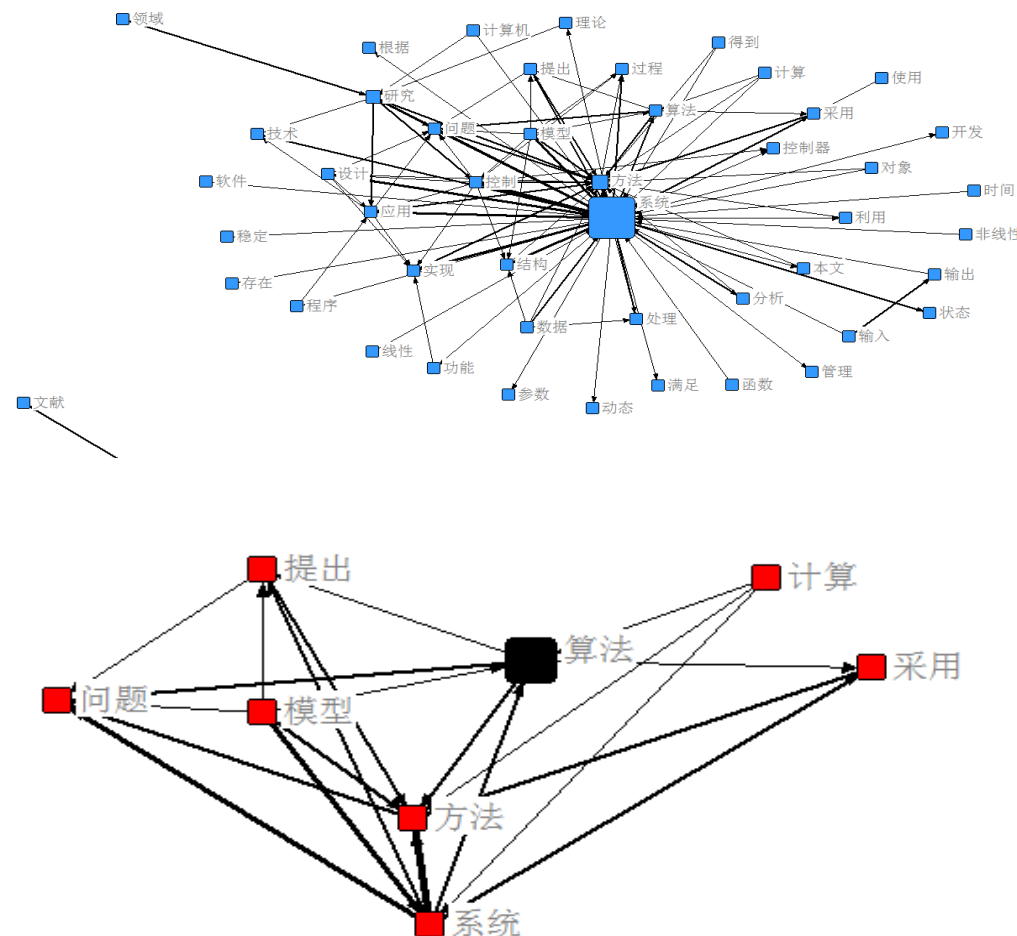
# 学科服务应用实例（一）



## ROSTCM可视化分析工具在计算机领域的应用

### 1997-1999年1358篇计算机行业新闻报道的热词表

词频	频率	特征词	TF-IDF 值	词频	频率	特征词	TF-IDF 值
系统	23483	接口	27	对象	6540	处理	10
控制	11547	数据库	22	研究	6428	提交	8
方法	10896	符号	21	过程	6279	生成	8
算法	10245	用户	16	实现	6066	语句	7
模型	9293	开发	15	函数	5626	输入	7
数据	8256	服务器	12	设计	5539	数据	7
问题	7745	生成器	12	程序	5478	网络	6
应用	7154	代码	11	计算	5146	访问	6
软件	6652	文件	10	处理	5088	功能	6

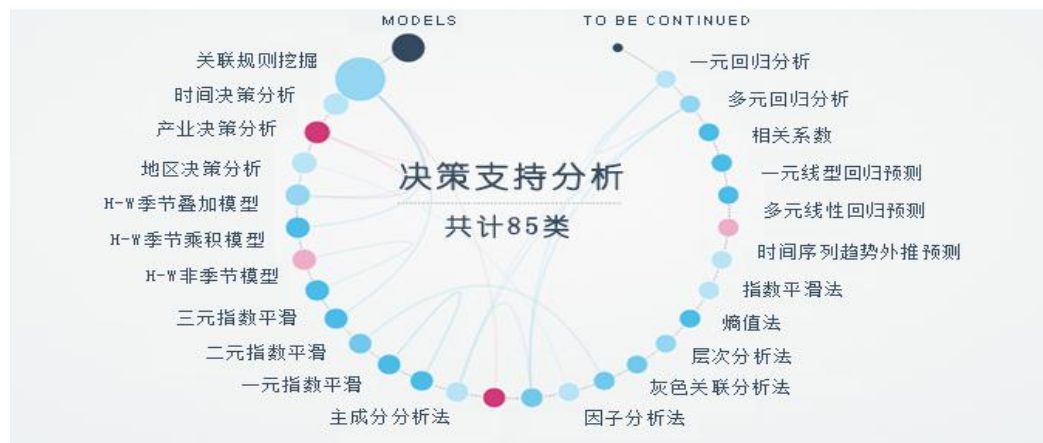


# 学科服务应用实例（二）



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 知网在线分析平台在学前教育中的应用



**决策支持研究模型**

相关性分析 | 统计预测 | **科学评价** | 决策模型

**选择建模方法**

熵值法 | **层次分析法 (AHP)** | 灰色关联分析法 | 因子分析法 | 主成分分析法

1 请输入评价目标

学前教育综合发展水平

2 请选择评价时间和地区

维度:  时间  地区

时间:  年度 [2013年]  季度 [2019年] [11月]

**已选状态**

研究模型: 科学评价  
建模方法: 层次分析法 (AHP)

评价目标: 学前教育综合发展水平  
评价时间: 2013年  
评价地区: 北京市; 天津市; 河北省; 山西省; 内蒙古; 辽宁省; 吉林省; 黑龙江; 上海市; 江苏省; 浙江省; 安徽省; 福建省; 江西省; 山东省; 河南省; 湖北省; 湖南省; 广东省; 广西; 海南省; 重庆市; 四川省; 贵州省; 云南省; 陕西省; 陕西省; 甘肃省; 青海省; 宁夏; 新疆;

评价面1: 教师规模 / 0.500  
指标: 每十万人人口幼儿园平均在园生数 / 1.000 / ...

评价面2: 办园条件 / 0.500  
指标: 幼儿园图书室 / 0.333 / 正向 / 标准样本变...  
幼儿园生活用房 / 0.333 / 正向 / 标准样本...  
幼儿园行政办公用房 / 0.333 / 正向 / 标准...

评价目标	得分	熵值法排名	评价目标	得分	熵值法排名
广东省	1.949917	1	辽宁省	1.262938	16
山东省	1.76562	2	云南省	1.246179	17
河南省	1.756952	3	贵州省	1.212011	18
江苏省	1.585972	4	重庆市	1.204304	19
浙江省	1.558342	5	新疆	1.199126	20
湖南省	1.52287	6	上海市	1.162519	21
河北省	1.472199	7	黑龙江	1.154851	22
四川省	1.463926	8	内蒙古	1.146969	23
江西省	1.416328	9	甘肃省	1.132137	24
福建省	1.380551	10	北京市	1.113209	25
广西	1.370539	11	吉林省	1.109518	26
陕西省	1.365939	12	海南省	1.105991	27
湖北省	1.351286	13	青海省	1.060106	28
安徽省	1.322438	14	宁夏	1.053453	29
山西省	1.267989	15	天津市	1.051327	30
辽宁省	1.262938	16	西藏	1.028149	31

## ■人类命运共同体视域下的科技查新工作对学科服务的创新参考

综合数据分析结果，提出图书馆科技查新工作提供学科服务创新的具体策略，实现图书馆学科服务创新的新路径。



学术论文1

学术论文2

学术论文3

## ■基于知网在线分析平台的区域学前教育发展水平评价

拓展数据分析工具在各学科领域中的实践应用，简化繁杂的实证过程，有效提升学术研究效率。

## ■基于ROSTCM和NetDraw的内容分析——以教育部科技查新站Z08数据为例



运用文本挖掘工具分析科技查新词频分布情况，了解地区医学学科发展状况，对比全国宏观发展趋势提出参考建议。



# 课题价值



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

结合词频分析为科研人员项目申报提供创新决策参考

通过课题延伸探讨高校医学查新站的智库功能



探索在线分析工具与文本挖掘工具在学科领域的实践应用



探索高校图书馆学科服务应用新模式新方法



结合共词网络图谱为科研管理者政策导向制定提供价值参考



# 研究不足



西安交通大学  
XI'AN JIAOTONG UNIVERSITY



选择偏好导致的  
样本量不全



分析软件选  
择相对局限



工具在学科中  
应用有待深化





西安交通大学  
XI'AN JIAOTONG UNIVERSITY

Thanks for everyone  
欢迎交流指导



QQ: 358304096