



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



应需而动的 资源建设创新模式

上海交通大学图书馆 李芳 黄镛

2015年5月12日



提纲

挑战和思考

实践与探索

- 自动分类专家系统

- 电子资源统计分析系统

结束语





上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

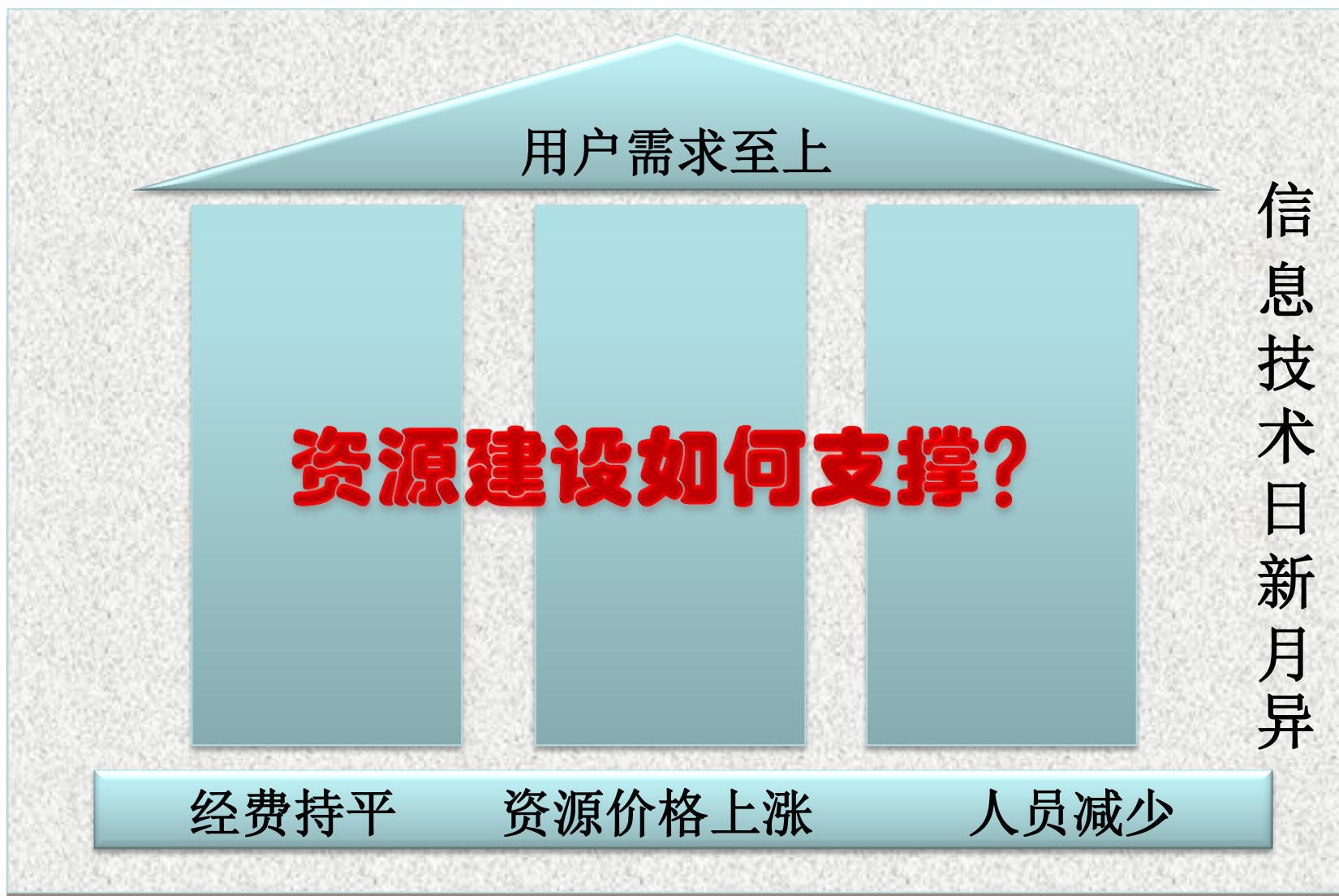


挑战与思考

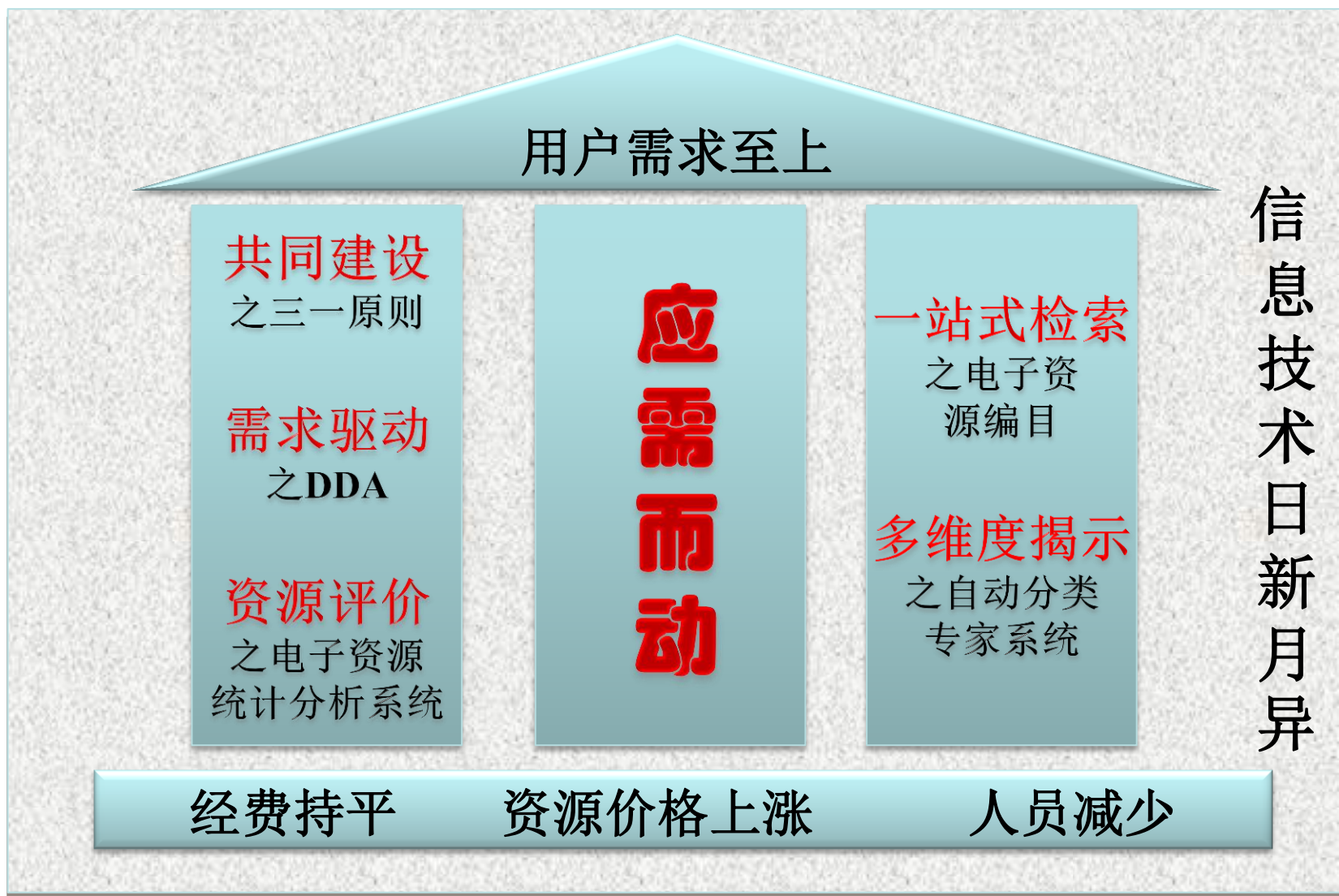




挑战和思考



挑战和思考





上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



实践与探索



自动分类专家系统—设计思路

设计目标

- 实现文献分类自动化，满足电子资源快速增长需求
- 构建国内外主流知识组织体系的映射系统

设计思路

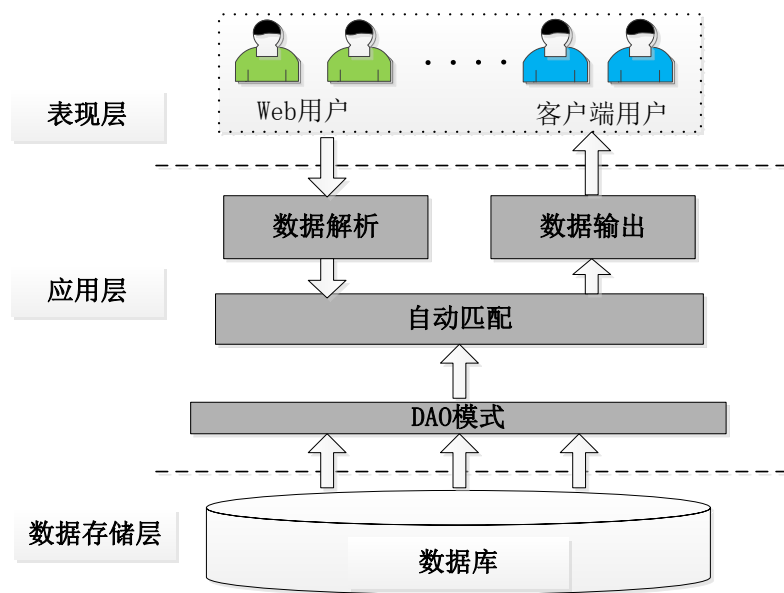
■ 覆盖四大知识组织体系

☞ DDC、LCC、LCSH、CLC

■ 基于同现映射的方法

■ 制定联合匹配规则

■ 搭建三层模式的系统结构



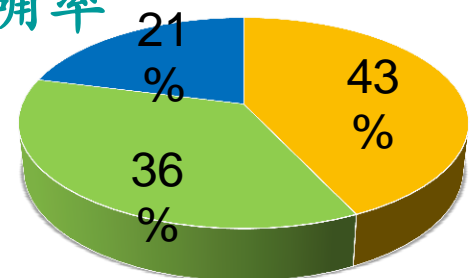
自动分类专家系统—数据收集

小样本测试

- 样本数量对映射结果有显著影响
- 引入LCSH和LCC，可以提升覆盖率和准确率

数据收集

- 共收集48万条数据，有效数据25万条
- 测试样本，1000条随机数据同时包含



■ 文科 ■ 理科 ■ 工科

CLC、DDC和LCSH字段

数据清理

- 特殊字符处理规则

字符	处理规则	所在字段	
\r\n	去除该字符	LCC, DDC, CLC, 650\$a	
/	去除该字符	LCC, DDC, CLC	
	保留该字符	650\$a	650\$a English language, Second language acquisition, EDUCATION / General HISTORY / Modern / 20th Century Fluid mechanics, Heat, SCIENCE / Mechanics / Dynamics / Thermodynamics Fluid mechanics, SCIENCE / Mechanics / General
.	去除该字符	DDC, CLC	

自动分类专家系统—构建映射表

映射表构建

- 对有效MARC记录分别抽取分类号和主题词，进行同现映射频次统计
- 对数据库中的隐藏的**关联规则**的挖掘，形成自动分类的基础，通过**支持度**、**置信度**筛选和构建

082_DDC	093_CLC	频次总计 (F1)	一对多总频次 (F2)	一对多统计 (N)	概率
620.11	TB3	1	7	4	1/7
620.11	TB301	4	7	4	4/7
620.11	TB381	1	7	4	1/7
620.11	TU5	1	7	4	1/7



自动分类专家系统-关键技术

机器统计学习方法

- 可分别根据DDC-CLC、LCSH-CLC、LCC-CLC对应数据统计和映射规则处理的需求，灵活设置各参数和阈值。

属性加权方法（Weighting Methods）

- LCSH与CLC相关度计算，呈**正相关关系**
- LCSH的排列顺序对匹配相关度的影响

聚类方法（Cluster Method）

- 上位类截取、上位类归并
- 完整的中图法三级类列表
- 特殊类目的提取



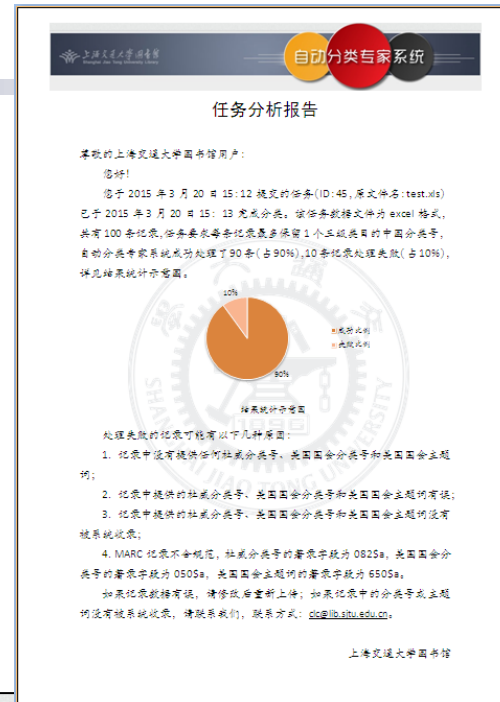
自动分类专家系统-成效

实现高效准确的自动分类

- 三级类目覆盖率100%
- 测试准确率90%
- 自动分类达200条记录/分钟
- 兼容不同数据格式

提升知识发现和服务效率

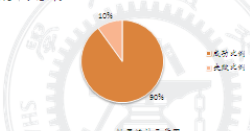
- 增加资源的发现途径
- 丰富资源揭示维度
- 精确统计学科资源
- 提升电子资源编目效率



尊敬的上海交通大学图书馆用户：

您好！

您于 2015 年 3 月 20 日 15:12 提交的任务(ID:45, 原名:任务: test.xls)已于 2015 年 3 月 20 日 15:13 完成分类。该任务数据文件为 excel 格式，共有 100 条记录，任务要求每行记录最多保留 1 个三级类目的中国分类号，自动分类专家系统成功处理了 90 条(占 90%)，10 条记录处理失败(占 10%)，详见结果统计示意图。



分类统计示意图

处理失败的记录可能有以下几种原因：

1. 记录中还没有提供任何社或分类号、美国国会分类号和英国国会主题词；
2. 记录中提供的社或分类号、美国国会分类号和英国国会主题词有误；
3. 记录中提供的社或分类号、美国国会分类号和英国国会主题词没有被系统收录；
4. MARC 记录不合规，社或分类号的著录字段为 0825a，美国国会分类号的著录字段为 0505a，英国国会主题词的著录字段为 6505a。

如果您的数据有误，请修改后重新上传。如果您的记录中的分类号或主题词没有被系统收录，请联系我们，联系方式：cs@lib.sjtu.edu.cn。

上海交通大学图书馆



自动分类专家系统

首页 上传数据 查看任务 用户信息 常见问题 [关于我们](#) [退出系统](#)

数据上传

查看任务

用户信息

常见问题

ERS-目标和功能

① 电子资源统计分析系统 (Electronic Resources Statistical Analysis System, 简称ERS)

② 主要目标

■ 有限经费向优势资源聚拢

- ⌘ 核心资源保障率分析
- ⌘ 需求分析

■ 优化学科资源结构

- ⌘ 基于学科的资源建设绩效评估

■ 辅助馆藏策略调整

- ⌘ 基于下载量、引用量的数据深度挖掘





ERS-文献计量指标和核心书目

外文期刊

- 文献计量指标: IF、SNIP、SJR
- JCR、ESI、Scopus、EI

外文图书

- Book Citation Index(BkCI)
- Choice Outstanding Academic Titles 2009-

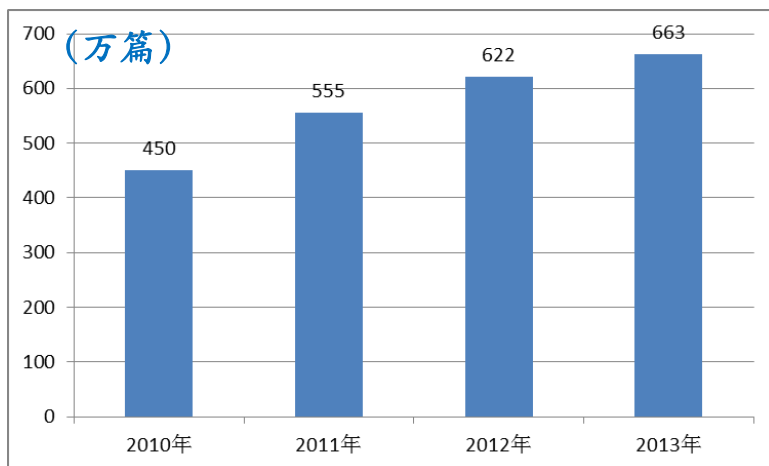
中文期刊

- 中文核心期刊要目总览
- 中国科技期刊引证报告
- 中文社会科学引文索引
- 中国科学引文数据库

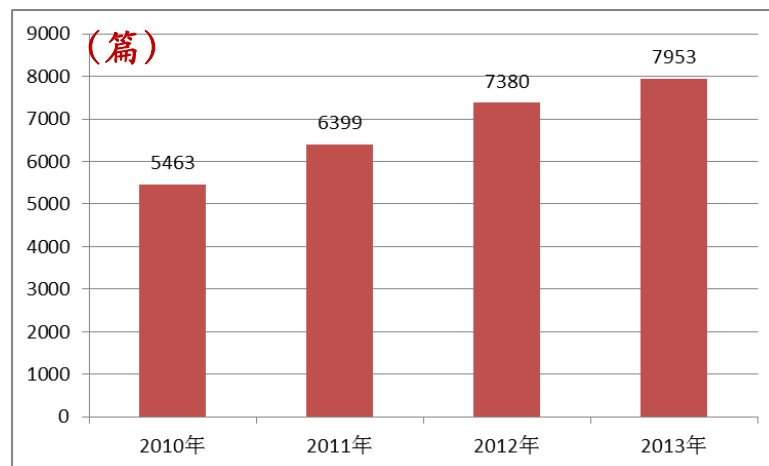


ERS-应用案例

- ④ 伴随着资源利用率的逐年上升，我校师生发文数量呈相应规律上升
- ④ 学科发展与文献利用率呈线性相关



2010-2013年外文电子资源全文下载量



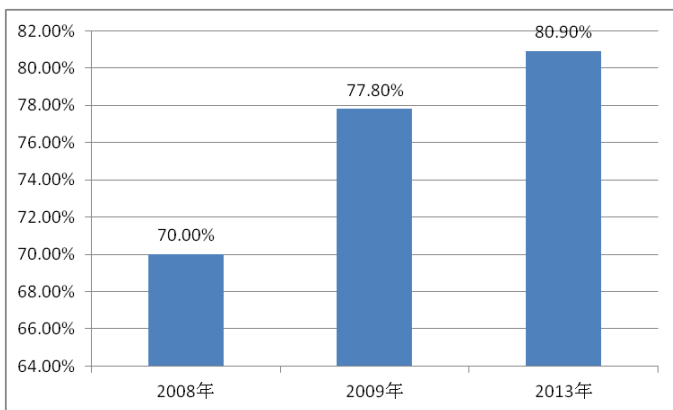
2010-2013年WoS收录的我校师生发文情况



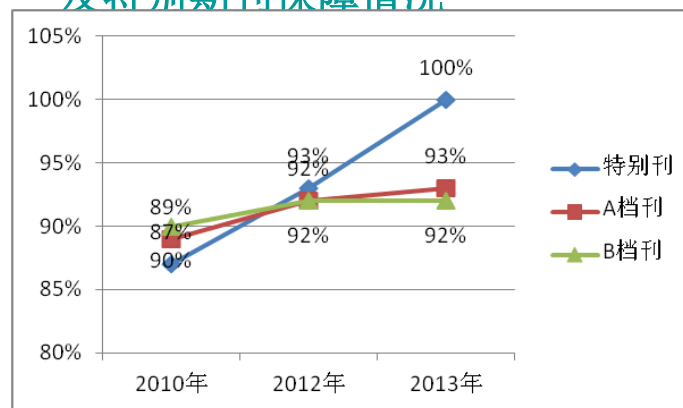
ERS-应用案例

上交大985三项资源建设效益

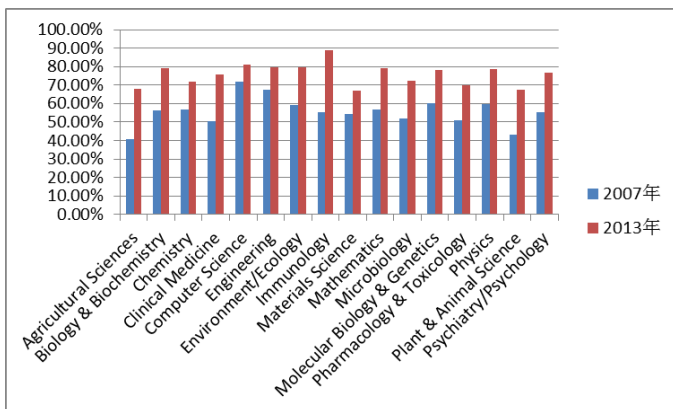
2008-2013年JCR来源刊保障情况



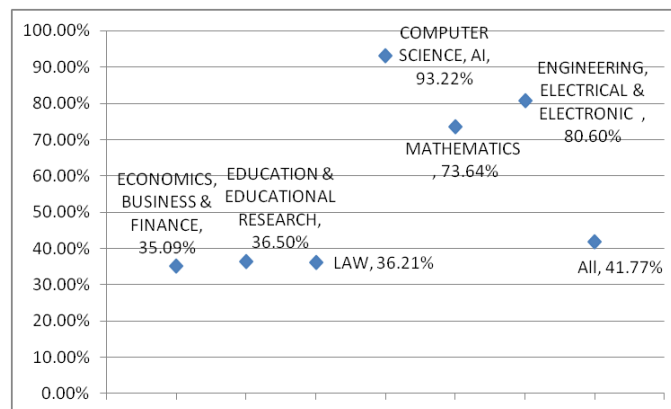
2010-2013年上海交通大学“A/B档”及特别期刊保障情况



2007-2013年ESI来源刊保障情况

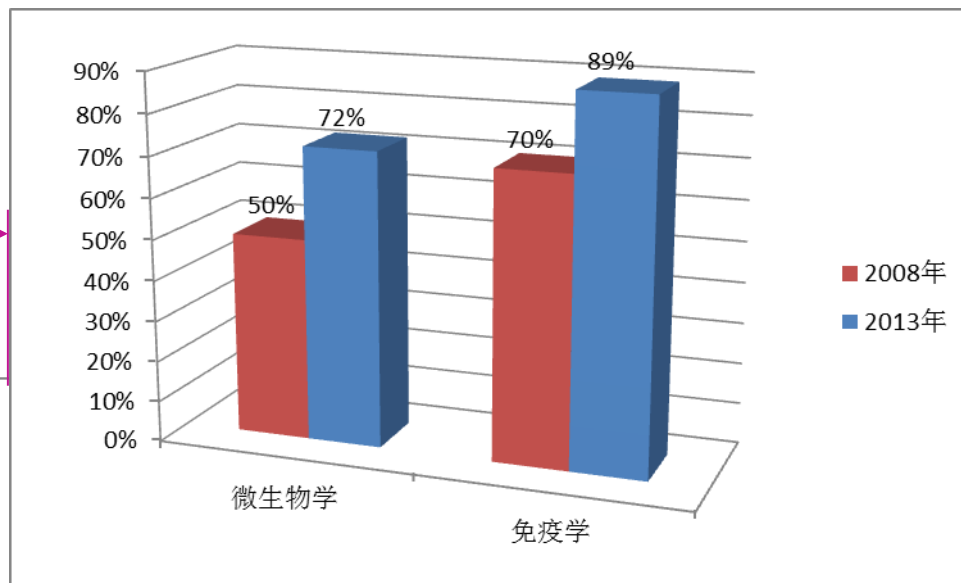
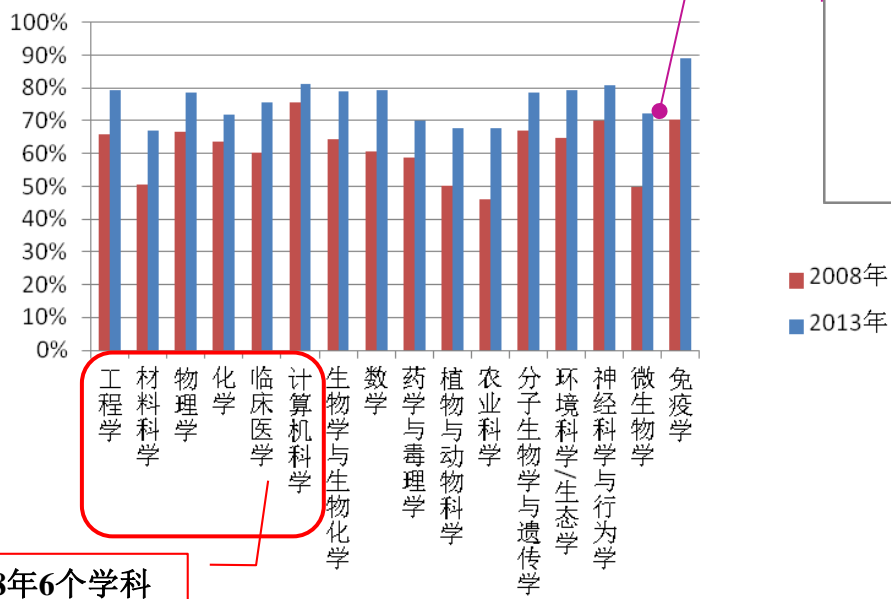


部分学科BkCI学术专著保障情况



ERS-应用案例

资源质量和成效提升显著



- 2008-2013年，我校ESI排名学科从6个增加到16个，所有新增机构的资源保障率均有较大幅度提升
- 例如：微生物学和免疫学的ESI来源刊资源保障率增幅达到约20%
- 资源保障率的提升为已有和新增ESI机构排名的学科提供了有力的支撑

2008-2013年我校进入ESI排名的16个学科刊保障情况

ERS-COUNTER报告收割和管理

- 外文期刊
- 外文图书
- 中文期刊
- 学科服务
- 外文期刊
- 外文图书
- 中文期刊
- Counter报告管理**
- COUNTER报告管理
- 维护Sushi协议
- 馆藏列表管理
- 订购数据库管理
- 引文列表管理
- 引文列表管理
- 引文分析

种类： 全部 电子期刊 电子图书 文摘事实

首字母： 全部 A B C D E F G H I J K L M N

数据库名称： 确定

任务列表

 ACM Digital Library	 AIP全文电子期刊及会议录数据库	 ACS数据库	 AGU美国地球物理学会数据库	 AIAA (美国航空航天学会) 全文电子期...
 APS全文电子期刊数据库	 ASTM SEDL	 Annual Reviews	 ASME数据库	 Begell Digital Libr...

为每个数据库建立文件夹，保存历史数据，方便备份

ERS-引文统计分析

引文统计比使用统计更精准、可靠，是文献保障支持科研的直接证据。

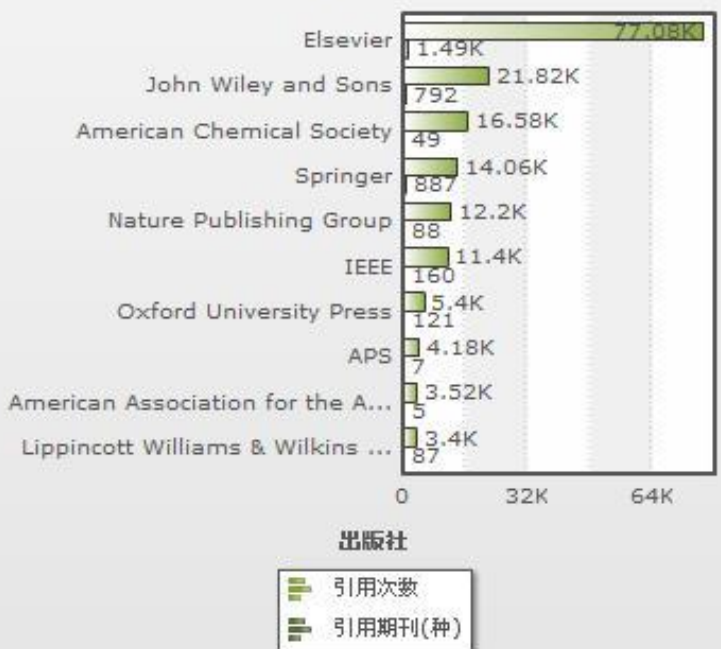
有引用出版社排名

Counter报告管理 引文分析

年份: **2013** 2014

数据库 出版社 年代分布 学科分布 语言分析 馆藏引用 导出报告

Top10 出版社



出版社	引用次数	期刊(种)
Elsevier	77083	1492
John Wiley and ...	21818	792
American Chemic...	16584	49
Springer	14058	887
Nature Publishi...	12203	88
IEEE	11397	160
Oxford Universi...	5397	121
APS	4178	7
American Associ...	3523	5
Lippincott Will...	3404	87



ERS_全文下载和引用量分级汇总

期刊利用符合二八法则，80%的利用来自20%的期刊。分级汇总帮助图书馆选“对”资源。

期刊名称	下载量
NATURE	3899
SCIENCE	3483
P NATL ACAD SCI USA	3102

年份： 2013 2014

下载量分级汇总	引用量分级汇总				↓ 下载
期刊年引用总量分级	引用量合计	引用量比率	期刊数量 (种)	期刊比率	
3000以上	10484	4.06%	3	0.01%	
2000--3999	9656	3.74%	4	0.01%	
1000-2999	17803	6.89%	14	0.03%	
600-999	19440	7.52%	26	0.06%	
400-599	21004	8.12%	43	0.09%	
100-399	82347	31.85%	444	0.96%	
50-99	34611	13.39%	497	1.07%	
10-49	46869	18.13%	2133	4.59%	
1-9	16299	6.30%	4822	10.39%	
0	0	0.00%	38445	82.80%	
合计	258513	100%	46431	100%	



ERS-期刊整体馆藏评估

ERS

电子资源统计分析系统
Electronic Resources Statistical Analysis System

测试账号
欢迎您

资源评估

外文期刊

外文图书

中文期刊

学科服务

外文期刊

外文图书

中文期刊

COUNTER报告管理

COUNTER报告管理

维护Sushi协议

馆藏列表管理

订购数据库管理

引文列表管理

引文列表管理

引文分析

馆藏期刊概览

资源评估

外文期刊

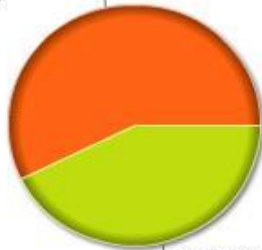
馆藏概览

概览

馆藏概览

馆藏概览

净刊量, 44.37K



重复刊量, 32.94K

净刊量 重复刊量

概览

名称	数量 (种)
馆藏总量 ?	77308
净刊量 ?	44371/57.40%
重复刊量 ?	32937/42.60%

核心期刊匹配率

JCR	ESI	SNIP	SJR
名称		数量 (种)	
JCR-2012 ?		10848	
净刊量 ?		44371	
匹配量 ?		8877/81.83%	

馆藏概览

JCR刊, 8.88K



非JCR刊, 1.97K

JCR学科分区	馆藏期刊	JCR合计	馆藏占比
No. 1	213	232	91.81%
Q1	4149	4548	91.23%
Q2	4098	4638	88.36%
Q3	3528	4128	85.47%
Q4	2956	4202	70.35%
去重共计	8877	10848	81.83%

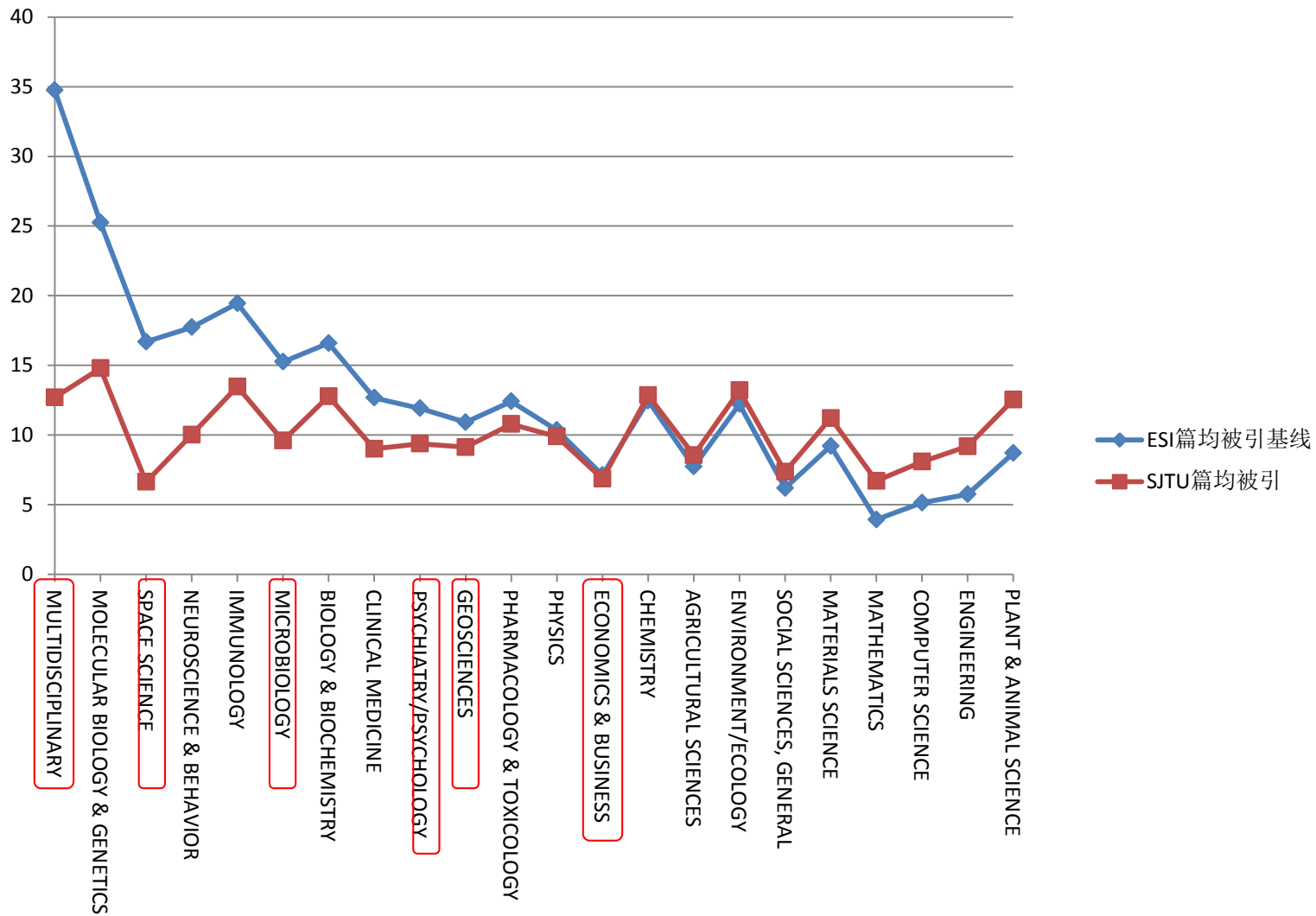


学科资源(下载—引用—馆藏保障)分析

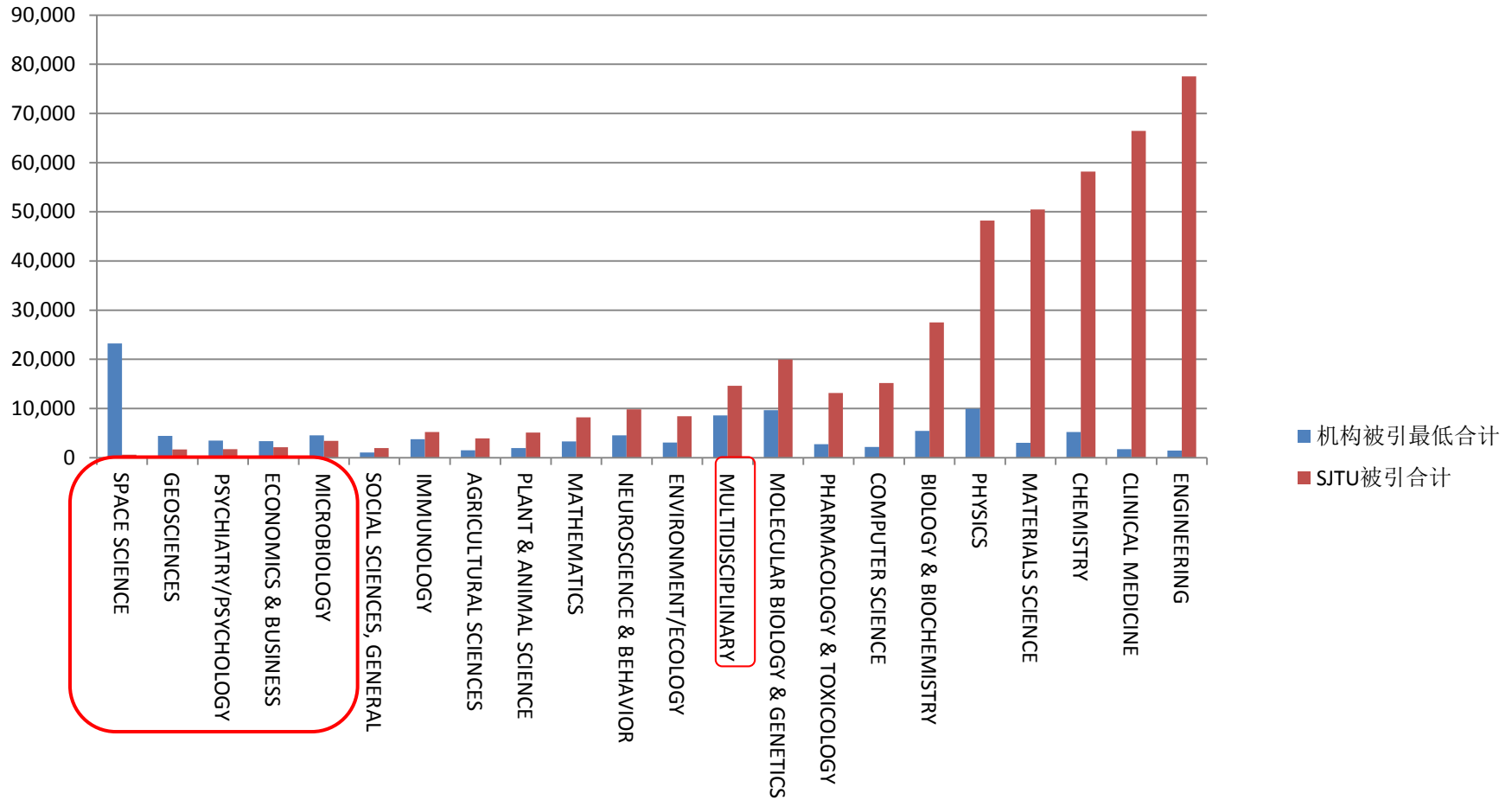
ESI	JCR	教育部学科门类	教育部一级学科				↓ 下载
	学科	有下载刊 (本)	下载量 (次)	有引用刊量 (本)	引用量 (次)	引用无馆藏 (本)	期刊保障率 ^②
	总计	14178	9368998	7569	318821	1599	76.87%
	哲学	95	10639	6	54	2	85.12%
	经济学	1405	102416	161	1872	2	95.42%
	法学	1007	58512	118	330	8	69.43%
	教育学	775	159058	269	2127	33	78.97%
	文学	558	20628	22	46	0	89.50%
	历史学	290	3226	1	1	0	94.80%
	理学	3124	3830335	2294	131533	483	72.02%
	工学	2375	2580230	1686	83958	282	70.52%
	农学	415	119254	170	2103	41	59.81%
	医学	2779	2199427	2315	78342	744	68.43%
	管理学	1261	285273	527	18455	4	81.39%
	艺术学	94	0	0	0	0	70.48%



ERS-学科篇均被引基线比较



ERS-学科Citations基线比较





ERS-发文、被引出版社分布

main Publisher	citations	发文	篇均被引	被引占比
Elsevier	203597	13895	14.65	35.75%
Wiley-Blackwell	39521	3270	12.09	6.94%
Springer	37434	6524	5.74	6.57%
ACS	29921	1103	27.13	5.25%
IEEE	25401	1705	14.90	4.46%
APS	13283	613	21.67	2.33%
Macmillan	13011	386	33.71	2.28%
RSC	10049	724	13.88	1.76%
IoP	9949	1114	8.93	1.75%
Science Press	9900	2899	3.41	1.74%
AIP	7791	654	11.91	1.37%
Taylor & Francis	7088	1319	5.37	1.24%
Wolters Kluwer Health	7025	670	10.49	1.23%
Shanghai Jiaotong University Press	6864	5617	1.22	1.21%
OUP	6675	547	12.20	1.17%
Public Library of Science	5098	781	6.53	0.90%
Optical Society of America	4892	433	11.30	0.86%
National Academy of Sciences	3169	72	44.01	0.56%
共计 (1385个出版社)	569486	75832	7.51	100.00%

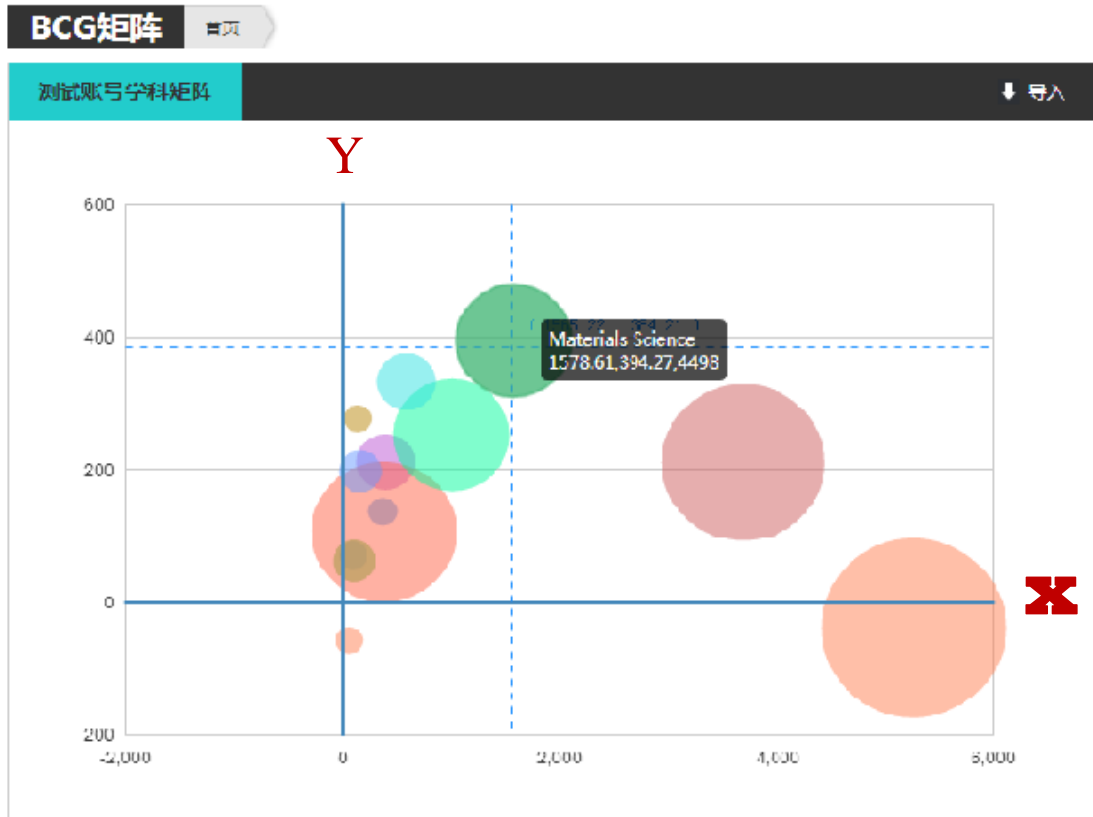
ERS-学科或资源发展决策-BCG矩阵



高

学科Citations Per Paper

低



Boston Matrix（市场增长率与相对市场份额矩阵）

- 用于机构绩效指标数据分析
- **X轴**：学科-机构Citations Per Paper基线
- **Y轴**：学科-机构Citations基线

Boston Matrix同样适用于数据库评估

学科Citations

高



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



结束语





结束语

技术发展的驱动力总是来自两个方向

- 更加满足需求
- 更加精致高效

探索一种全新的文献资源建设模式

- 更加契合用户需求
- 确保提升资源利用率
- 使有限的经费发挥更大的作用
- 维持馆藏资源结构协调、系统地发展





上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



谢谢!

