

REUTERS/Pawel Kopczynski

# 如何利用情报分析工具助力高校图书馆开展深层次学科服务

---

## ——TDA培训课程（一）

段鑫龙

产品与解决方案部

汤森路透

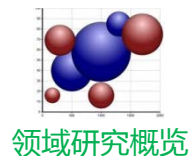


THOMSON REUTERS



# Thomson Data Analyzer

- 基于VantagePoint 技术的情报分析工具；
- “离线” 工具。它本身并不包含任何数据，也不检索任何数据库



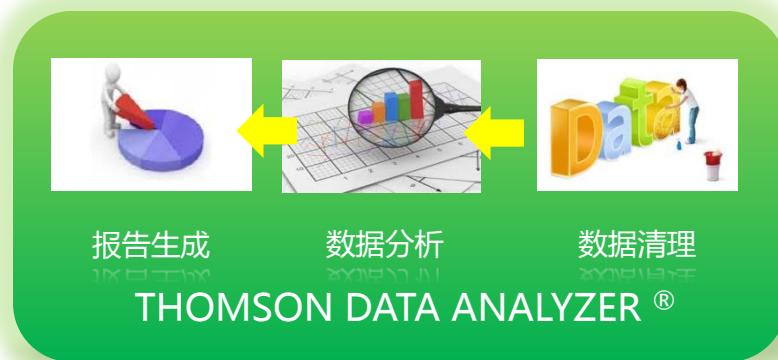
领域研究概览



智能分析报告



合作分析展现



## TDA已有的过滤器

Thomson Innovation - Patents (TDA format)
STN - DPCI (Text (indented))
STN - DPCI (Text)
STN - DWPI (Text (indented))
STN - DWPI (Text)
Thomson Innovation - Business (Tagged Text)
Thomson Innovation - Literature (TDA format)
WoK - DII (Field-Tagged text)
WoK - INSPEC (Field-Tagged text)
WoK - WoS (Field-Tagged text)





# Thomson Data Analyzer

---

- 1. TDA主要功能**
- 2. TDA分析实例**
- 3. TDA常见问题 Q&A**

# TDA的主要功能

## 1.数据导入

- |                |
|----------------|
| 1.1 结构化数据都可导入  |
| 1.2 添加/融合字段    |
| 1.3 数据集/记录融合   |
| 1.4 重复项（去重/合并） |

## 2.数据清理

- |          |
|----------|
| 2.1 建立词表 |
| 2.2 修改词表 |
| 2.3 使用词表 |

## 3.数据分析

- |                                 |
|---------------------------------|
| 3.1 各种脚本，比如线图、饼图、柱状、aduna、聚类圆群等 |
| 3.2 技术报告、公司报告、公司比较等脚本           |
| 3.3 矩阵                          |
| 3.4 图谱                          |
| 3.5 词频逆文档(TF-IDF)               |

## 4.数据导出

- |             |
|-------------|
| 4.1 导出字段化记录 |
| 4.2 导出原始记录  |

# TDA的主要功能

## 1.数据导入

- |                |
|----------------|
| 1.1 多种格式数据导入   |
| 1.2 添加/融合字段    |
| 1.3 数据集/记录融合   |
| 1.4 重复项（去重/合并） |

## 2.数据清理

- |          |
|----------|
| 2.1 建立词表 |
| 2.2 修改词表 |
| 2.3 使用词表 |

## 3.数据分析

- |                                 |
|---------------------------------|
| 3.1 各种脚本，比如线图、饼图、柱状、aduna、聚类圆群等 |
| 3.2 技术报告、公司报告、公司比较等脚本           |
| 3.3 矩阵                          |
| 3.4 图谱                          |
| 3.5 词频逆文档(TF-IDF)               |

## 4.数据导出

- |             |
|-------------|
| 4.1 导出字段化记录 |
| 4.2 导出原始记录  |

# TDA的主要功能

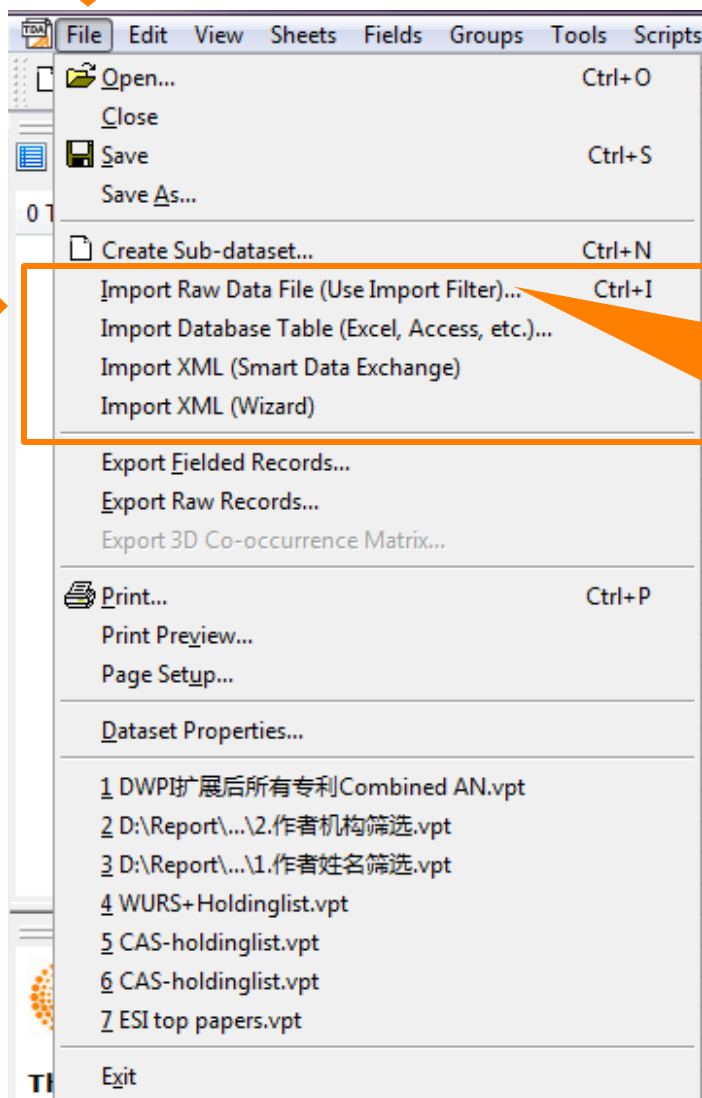
## 1. 数据导入

1.1 结构化数据均可导入

1.2 添加/融合字段

1.3 数据集/记录融合

1.4 重复项（去重/合并）



Raw data需要过滤器，目前TDA已有的过滤器是WOS/TI/Inspec，如果想导入其他平台的数据到TDA中，需要利用Tools->Import Filter Editor

# TDA的主要功能

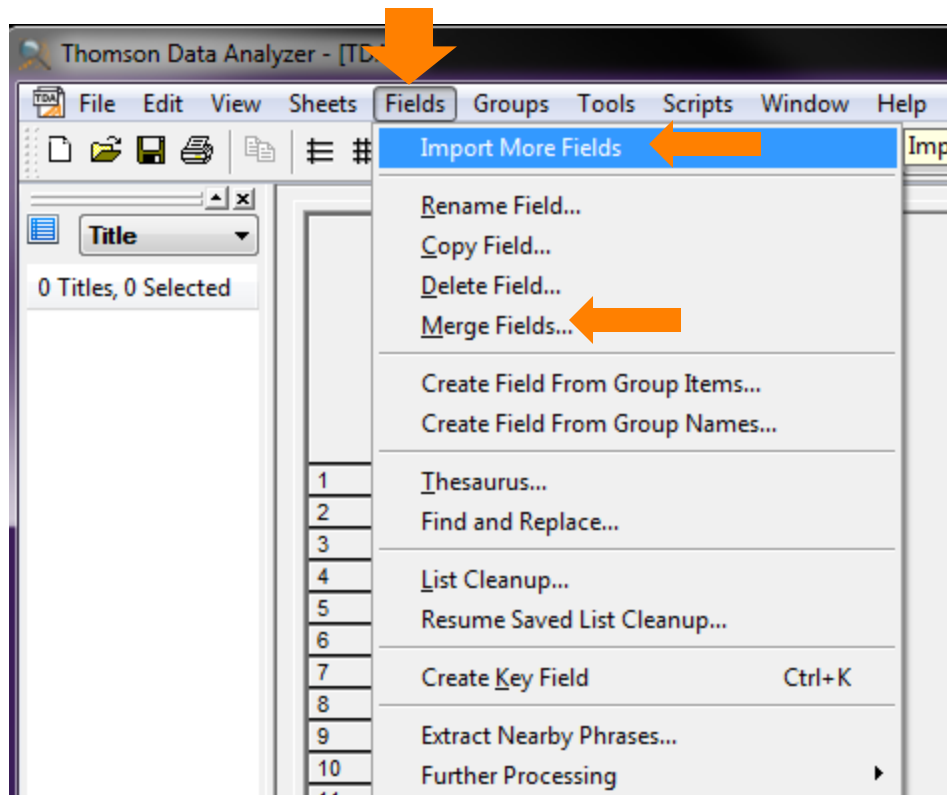
## 1. 数据导入

1.1 结构化数据均可导入

1.2 添加/融合字段

1.3 数据集/记录融合

1.4 重复项（去重/合并）





# 两个字段中项与项的融合

## Utility-concatenate Fields

SUMMARY SHEET		Number of R
Field	Number of Items	
(filters)	>=500	
1st author's	281	100%
Author Affiliations (1st)	42	
Author Affiliations (1st) (Full)	169	
Author Affiliations (Authors - Organization)	653	99%
Author		
Organization		
Author Affiliations (City and Country)	134	100%
Author Affiliations (Full)	484	100%
Author Affiliations (Organization and City and Country)	234	100%
Author Affiliations (Organization Only)	218	
Authors	1,168	
Authors (1st)	259	100%
Authors (1st) (Full Name)		
Authors (1st): Author Affiliations (1st)		
Authors (Full Name)		
Authors (Initials)		
Cited Authors		
Cited Journal		

- Report - World Map
- Report - Plot Classification
- Report - Browse Classification Tree
- Export - Copy Publication Numbers to Clipboard
- Utility - AND search
- Utility - Concatenate Fields**
- Utility - Create All Lists
- Utility - Combine Groups
- Utility - Make These
- Utility - Mark Un

Concatenate two fields  
The second field can

② 利用Utility-concatenate Fields

① 选择待融合的两个字段

③ 融合第一作者单位和第一作者全名之后

	# Records	# Instances	1st author's
1	3	3	Beijing Forestry Univ, MOE Key Lab Wooden Mat Sci & Applicat, Beijing 100083, Peoples R China Lu, Yi
2	2	2	Beijing Forestry Univ, Beijing Key Lab Lignocellulos Chem, Beijing 100083, Peoples R China Dai, L-j
3	2	2	Beijing Forestry Univ, Coll Biol Sci & Technol, Natl Engrn Lab Tree Breeding, Beijing 100083, Peoples R China Chen, Jinhui
4	2	2	Beijing Forestry Univ, Coll Environm Sci & Engrn, Beijing 100083, Peoples R China Chen, Xuelin
5	2	2	Beijing Forestry Univ, Coll Landscape Architecture, Beijing 100083, Peoples R China Hong, Yan
6	2	2	Beijing Forestry Univ, Coll Mat Sci & Technol, Beijing 100083, Peoples R China Bai, Haolong
7	2	2	Beijing Forestry Univ, Inst Microbiol, Beijing 100083, Peoples R China Han, Mei-Ling
8	2	2	Beijing Forestry Univ, MOE Key Lab Wooden Mat Sci & Applicat, Beijing 100083, Peoples R China Peng, Yao
9	2	2	Beijing Forestry Univ, MOE Key Lab Wooden Mat Sci & Applicat, Beijing 100083, Peoples R China Zhang, Jizhi
10	2	2	Renmin Univ China, Dept Chem, Beijing 100872, Peoples R China Sun, Xiaofu
11	2	2	Tsinghua Univ, Ctr Earth Syst Sci, Minist Educ, Key Lab Earth Syst Modeling, Beijing 100084, Peoples R China Yang, Jun
12	1	1	Beihang Univ, Sch Instrumentat Sci & Optoelect Engrn, Beijing 100083, Peoples R China Peng, Cong



# TDA的主要功能

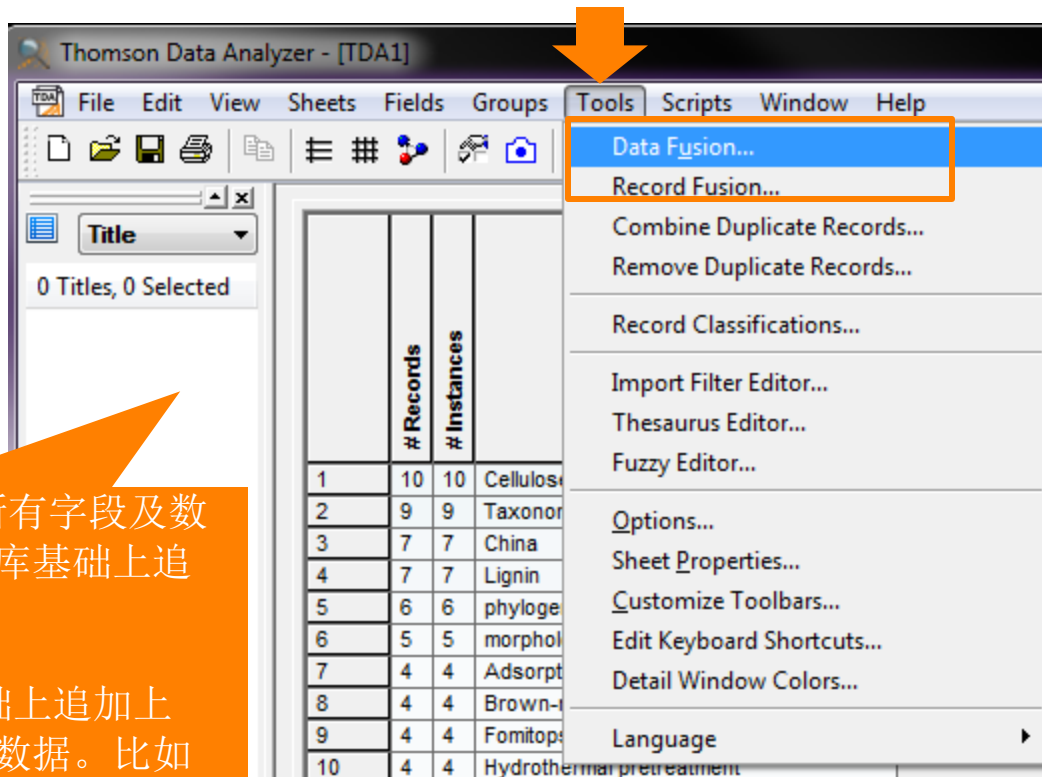
## 1. 数据导入

1.1 结构化数据均可导入

1.2 添加/融合字段

1.3 数据集/记录融合

1.4 重复项（去重/合并）



**Data Fusion:** 是两个数据集中所有字段及数据的融合，比如在历史机构知识库基础上追加上新年度的数据；

**Record Fusion:** 是在主集合基础上追加另外另一个文档中的某几个字段及数据。比如在SCI机构知识库中加入期刊影响因子信息

# TDA的主要功能

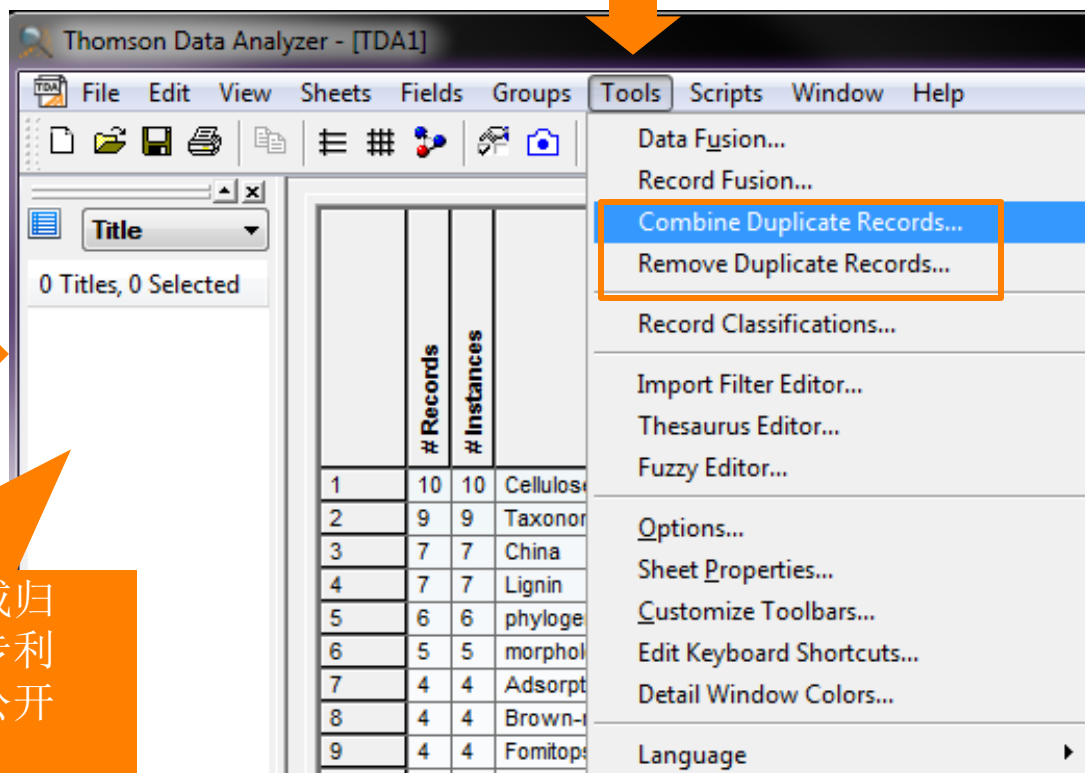
## 1. 数据导入

1.1 结构化数据均可导入

1.2 添加/融合字段

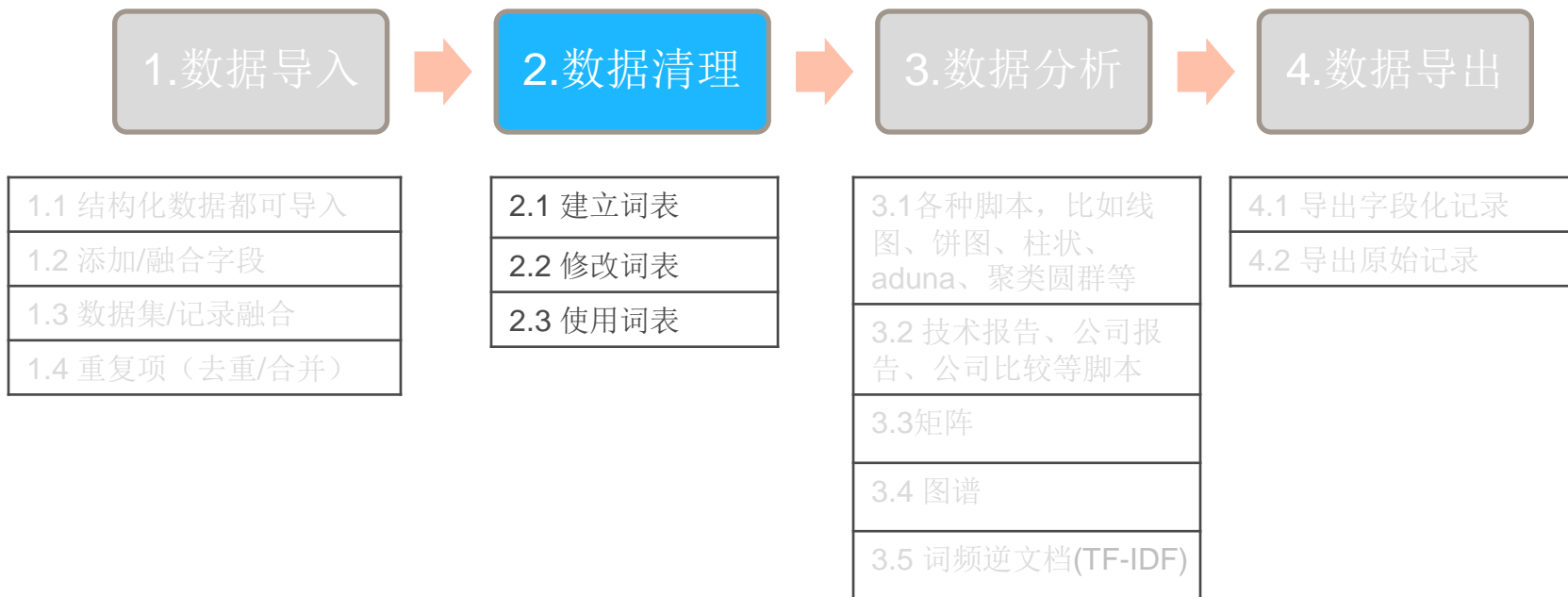
1.3 数据集/记录融合

1.4 重复项（去重/合并）



涉及到归并去重时可利用去重或归并功能。比如专利数据，一件专利会对应多个文本：申请文本、公开文本、授权文本等。

# TDA的主要功能



# TDA的主要功能

---

## 2. 数据清理

### 2.1 建立词表

### 2.2 修改词表

### 2.3 使用词表

三种方法：

其一，自动清洗：Fields->list cleanup，比如发明人、专利权人、文本字段

其二，查找分组进行标引：

比如，如何提炼成立本校各学院SCI知识库，以北京林业大学为例。大致思路：

通常会利用author affiliation (full) 字段，在其中搜索各学院然后添加到不同的组，但因为author affiliation (full) 中还有其他机构，这给搜索学院带来噪声。所以，最好首先挑出北京林业大学的项，单独添加到组后，再单独将这些项成立一条新字段

(Fields->Create field From Group Items)。之后把各学院变体放到不同组后，可利用组成立叙词表 (Group->Create Thesaurus Using Groups)

其三，对现有的excel表加工编写成词表 ( Scripts->Utility-Make Thesaurus) 。



# TDA的主要功能

---

## 2. 数据清理

2.1 建立词表

2.2 修改词表

2.3 使用词表

对已保存的词表进行修改，有两种方法：

其一，在词表编辑器中打开词表进行修改（Tools->Thesaurus Editor）；

其二，直接在TDA系统文件中，在thesaurus文件夹中直接打开之前保存的词表进行编辑。

第一种方法适合少量修改；第二种方法适合批量修改

# TDA的主要功能

## 2. 数据清理

2.1 建立词表

2.2 修改词表

2.3 使用词表

如何利用之前建好的叙词表？

两种方式：

其一，直接替换字段下各项（Fields->Thesaurus）；

其二，以组的形式呈现（Groups->Group Using Thesaurus）

需要注意：

一，词表在用之前一定要确保匹配规则（精确匹配、前方一致、后方一致还是包含）；

二，待匹配的字段必须要滤除噪声。比如如果拿北京林业大学只包含各学院的变体词表去清理author affiliation（full）极可能命中的物理学院有其他学校的物理学院。



# 以组的形式呈现 (Groups->Group Using Thesaurus)

以组形式匹配

	记录数量	#频次	Author Affiliations (Full)
1	252	271	Sch Mat Sci & Engr
2	251	253	Dept Phys
3	44	47	Ctr High Energy Phys
4	34	34	Dept Mat Sci & Engr
5	3	3	Coll Mat Sci & Engr
6	2	2	Phys Dept
7	1	1	Center High Energy Phys
8	1	1	Phys Dept & Tsinghua Ctr Astrophys
9	1	1	Phys Dept & Tsinghua Ctr Astrophys THC
10	287	292	Dept Elect Engr
11	207	212	Sch Environm
12	194	195	Dept Chem
13	162	177	Grad Sch Shenzhen
14	144	165	Sch Life Sci
15	139	145	Sch Med
16	134	137	Dept Automat
17	116	119	Dept Chem Engr
18	101	104	Dept Engr Mech



	记录数量	#频次	Author Affiliations (Full)	ctr high Energy Phys	Sch Mat Sci Engr	Dept Phys
1	252	271	Sch Mat Sci & Engr	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2	251	253	Dept Phys	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3	44	47	Ctr High Energy Phys	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	34	34	Dept Mat Sci & Engr	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
5	3	3	Coll Mat Sci & Engr	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6	2	2	Phys Dept	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
7	1	1	Center High Energy Phys	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	1	1	Phys Dept & Tsinghua Ctr Astrophys	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
9	1	1	Phys Dept & Tsinghua Ctr Astrophys THC	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
10	287	292	Dept Elect Engr	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	207	212	Sch Environm	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12	194	195	Dept Chem	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	162	177	Grad Sch Shenzhen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14	144	165	Sch Life Sci	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15	139	145	Sch Med	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16	134	137	Dept Automat	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17	116	119	Dept Chem Engr	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18	101	104	Dept Engr Mech	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19	100	100	Dept Thermal Engr	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



# 直接替换字段下各项 (Fields->Thesaurus)

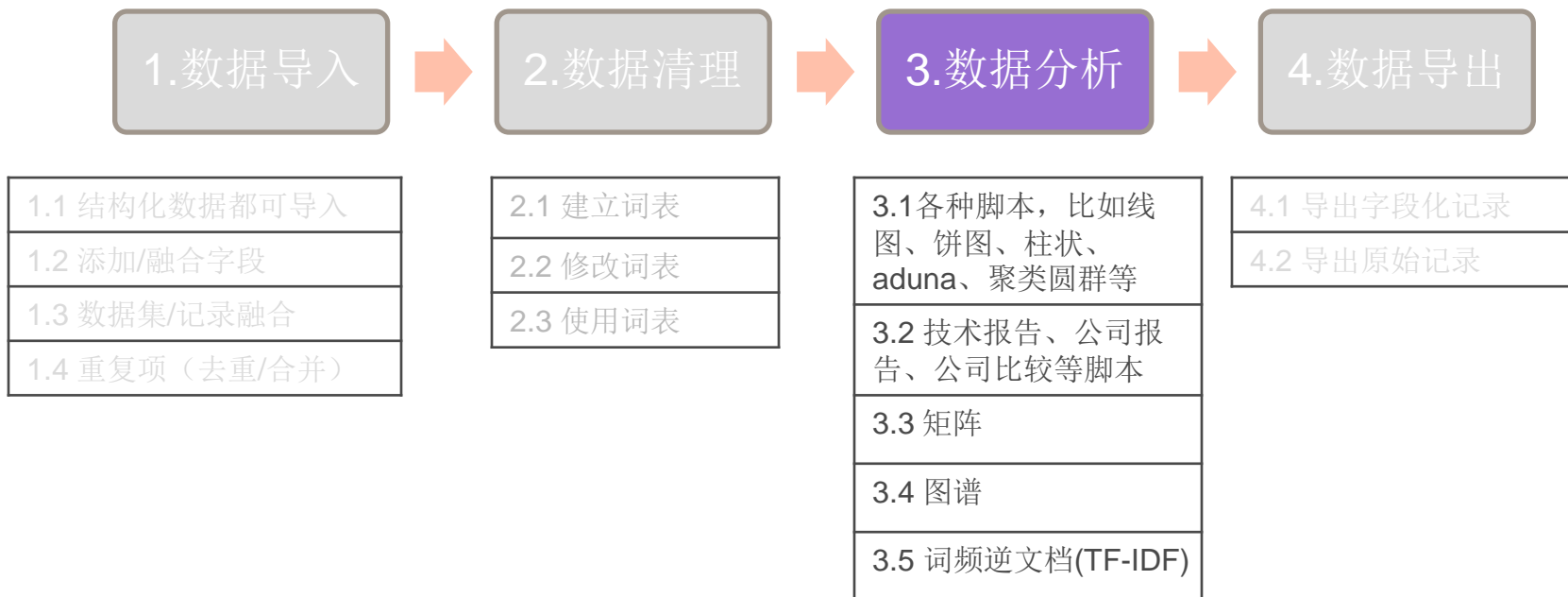
	记录数量	#频次	Author Affiliations (Full)
1	252	271	Sch Mat Sci & Engn
2	251	253	Dept Phys
3	44	47	Ctr High Energy Phys
4	34	34	Dept Mat Sci & Engn
5	3	3	Coll Mat Sci & Engn
6	2	2	Phys Dept
7	1	1	Center High Energy Phys
8	1	1	Phys Dept & Tsinghua Ctr Astrophys
9	1	1	Phys Dept & Tsinghua Ctr Astrophys THC
10	287	292	Dept Elect Engn
11	207	212	Sch Environm
12	194	195	Dept Chem
13	162	177	Grad Sch Shenzhen
14	144	165	Sch Life Sci
15	139	145	Sch Med
16	134	137	Dept Automat
17	116	119	Dept Chem Engn
18	101	104	Dept Engn Mech



	记录数量	#频次	Author Affilia ll):tsingl
1	289	308	Sch Mat Sci Engn
2	255	257	Dept Phys
3	45	48	ctr high Energy Phys
4	287	292	Dept Elect Engn
5	207	212	Sch Environm
6	194	195	Dept Chem
7	162	177	Grad Sch Shenzhen
8	144	165	Sch Life Sci
9	139	145	Sch Med
10	134	137	Dept Automat
11	116	119	Dept Chem Engn
12	101	104	Dept Engn Mech
13	100	100	Dept Thermal Engn
14	97	99	Dept Mech Engn
15	97	97	State Key Lab New Ceram & Fine Proc
16	90	90	Beijing 100084
17	82	82	Dept Comp Sci & Technol

直接改写字段项，可见记录数有变化

# TDA的主要功能



# TDA的主要功能

## 3. 数据分析

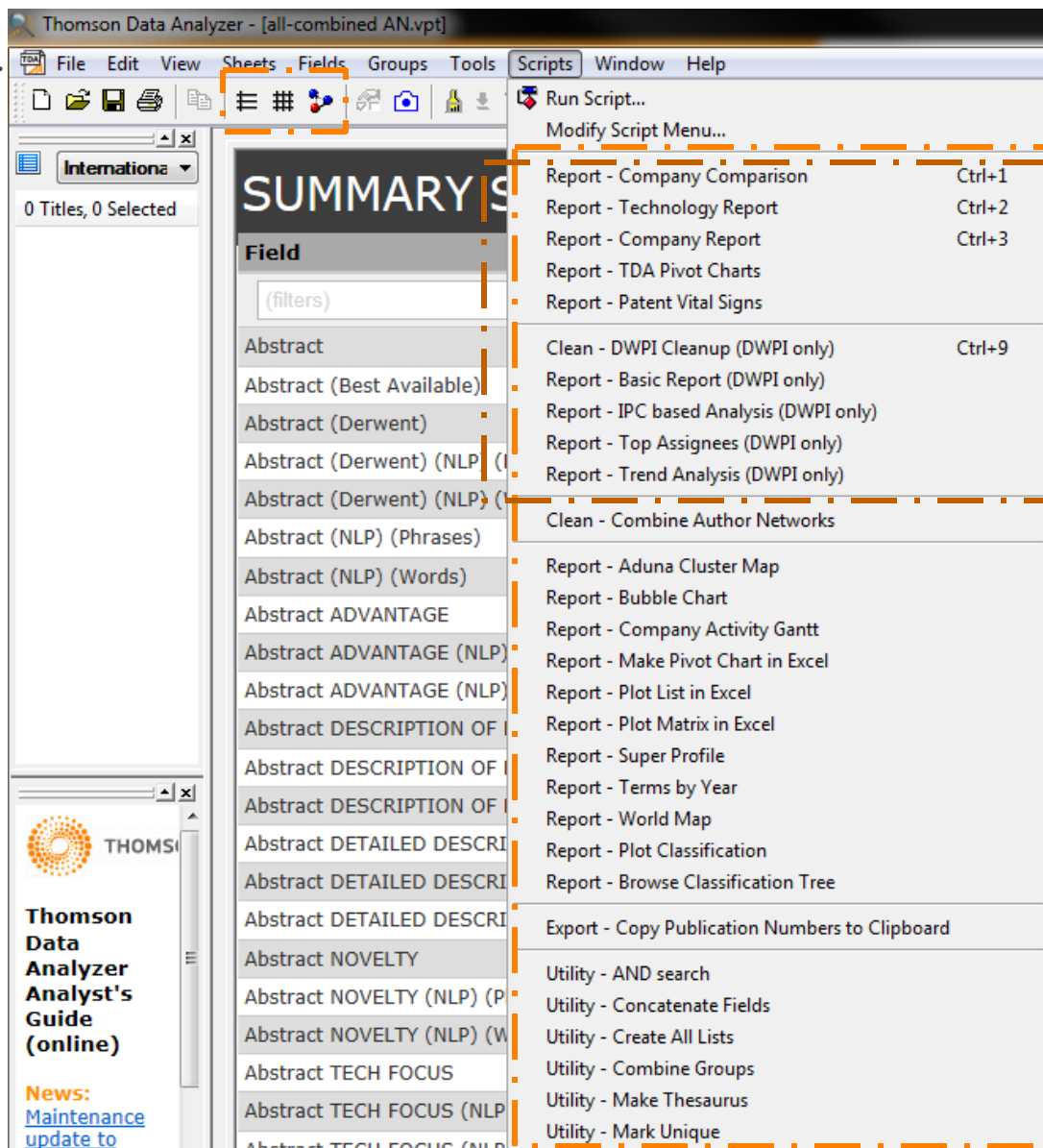
3.1 各种脚本，比如线图、饼图、柱状、aduna、聚类圆群等

3.2 技术报告、公司报告、公司比较等脚本

3.3 矩阵/矩阵可视化

3.4 图谱

3.5 词频逆文档



# TDA的主要功能

## 3.数据分析

3.1 各种脚本，比如线图、饼图、柱状、aduna、聚类圆群等

3.2 技术报告、公司报告、公司比较等脚本

3.3 矩阵/矩阵可视化

3.4 图谱

3.5 词频逆文档(TF-IDF)

可以解决哪些问题：

- ①与excel兼容，直接生成柱状图、线图、饼图、线图、泡泡图等；
- ②看合作（个人/机构等）
- ③直接生成技术报告、公司报告、公司比较报告；
- ④专利健康报告：patent vital signs

矩阵有共现、自相关、互相关；

- ① 共现：比如看各学院历年的发文量，他们在最近几年的活跃度；各专利权人专利的全球布局
- ② 自相关：看已有专利权人之间的合作；
- ③ 互相关：看哪些人正在做着相似的事；图谱中的自相关、互相关与矩阵原理相同

# TDA的主要功能

## 3. 数据分析

3.1 各种脚本，比如线图、饼图、柱状、aduna、聚类圆群等

3.2 技术报告、公司报告、公司比较等脚本

3.3 矩阵/矩阵可视化

3.4 图谱

3.5 词频逆文档(TF-IDF)



Reset		Keywords (author's)	1
		# Records	872
Key	# Records	Show Values $\geq 0.00$ and $\leq 7.42$ TFIDF (sqrt)	All Records
1	21	China	7.42
2	21	Lignin	7.42
3	17	Cellulose	7.05
4	13	Populus euphratica	6.59
5	13	taxonomy	6.59
6	12	poplar	6.45
7	11	Hemicelluloses	6.30
8	10	Microstructure	6.14
9	10	NaCl	6.14
10	10	Photosynthesis	6.14
11	9	Adsorption	5.96
12	9	evapotranspiration	5.96
13	9	Immunohistochemistry	5.96
14	9	phylogeny	5.96
15	8	Drought	5.76
16	8	Hydrothermal	5.76
17	8	Microwave	5.76
18	7	Climate change	5.54
19	7	Gene expression	5.54
20	7	inhibin	5.54
21	7	Morphology	5.54
22	7	Nanocomposites	5.54
23	7	Nanomaterials	5.54
24	6	Fractionation	5.30
25	6	functional mapping	5.30
26	6	Loess Plateau	5.30

词频逆文档 TF-IDF

**(Ctrl+T)启动!**

TF: Term Frequency;

IDF: Inverse Document Frequency

用统计学语言表达，就是在词频的基础上，要对每个词分配一个“重要性”权重。比如关于中国蜜蜂养殖，最常见的词（“的”、“是”、“在”）给予最小的权重，较常见的词（“中国”）给予较小的权重，较少见的词（“蜜蜂”、“养殖”）给予较大的权重。这个权重叫做“逆文档频率”（Inverse Document Frequency，缩写为IDF），它的大小与一个词的常见程度成反比。



# Thomson Data Analyzer

---

1. TDA主要功能
2. TDA分析实例
3. TDA常见问题 Q&A

# 基于科技文献的团队科研绩效分析

---

I 科技论文发展概况

II 科研合作

III 学术质量分析

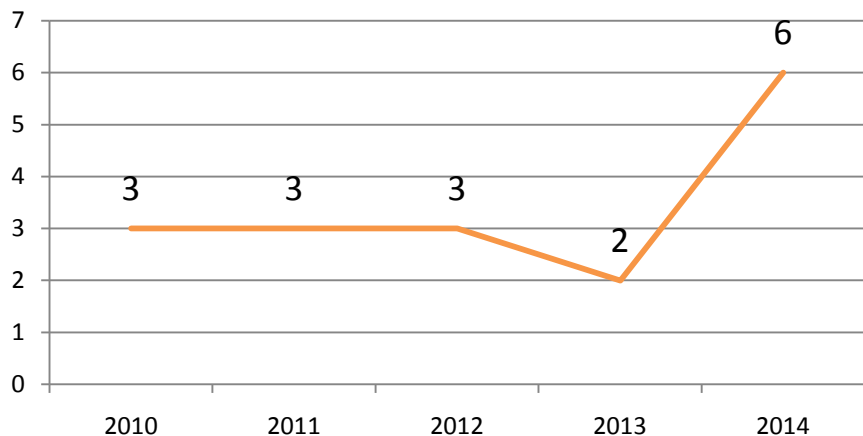
IV 对标分析





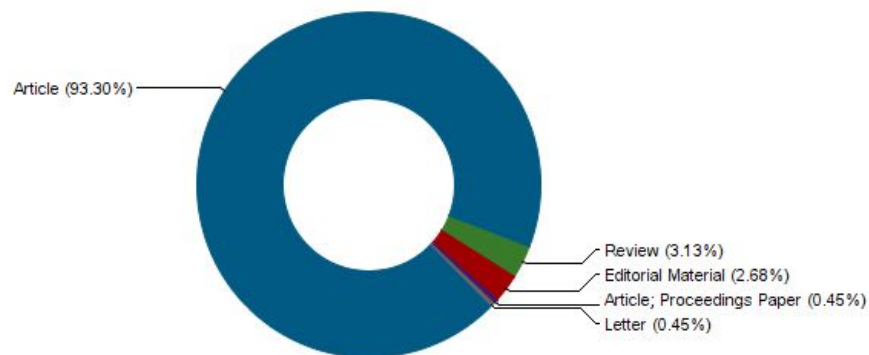
# I 科技论文发展概况

## 某团队文章发表趋势（单位：篇）



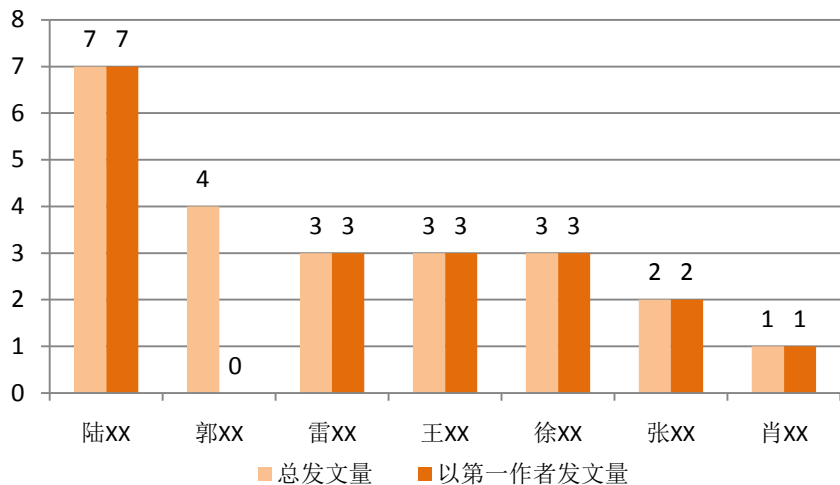
Report-Plot list in Excel

## 文献类型分布



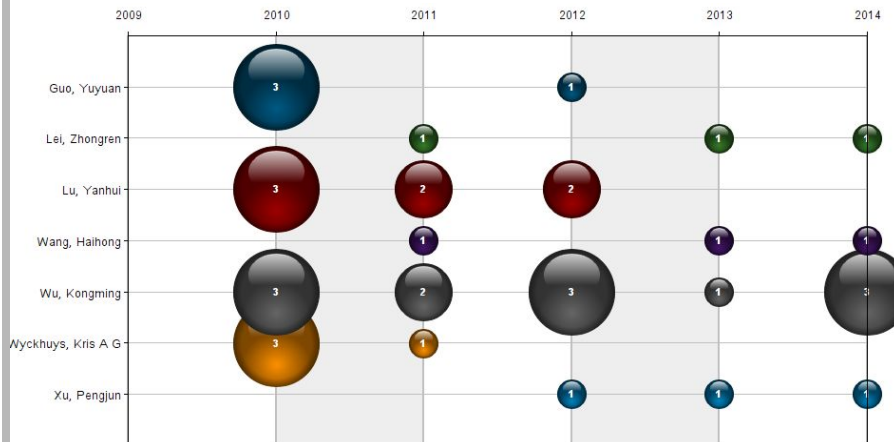
Report-Circle Chart

## 高发文作者及其第一作者文章情况



Report-Plot list in Excel

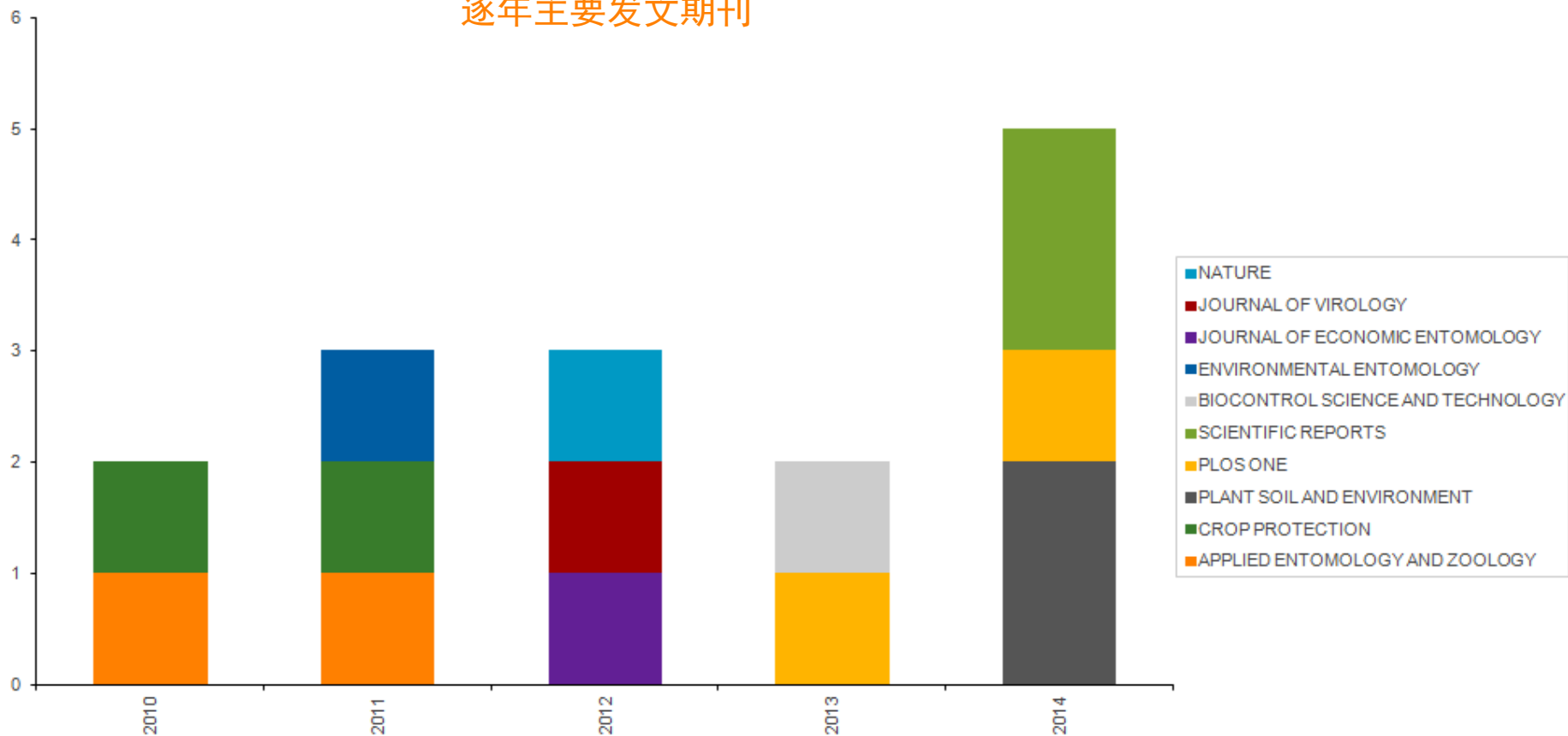
## 高发文作者2009-2014年间活跃度



Report-Bubble Chart

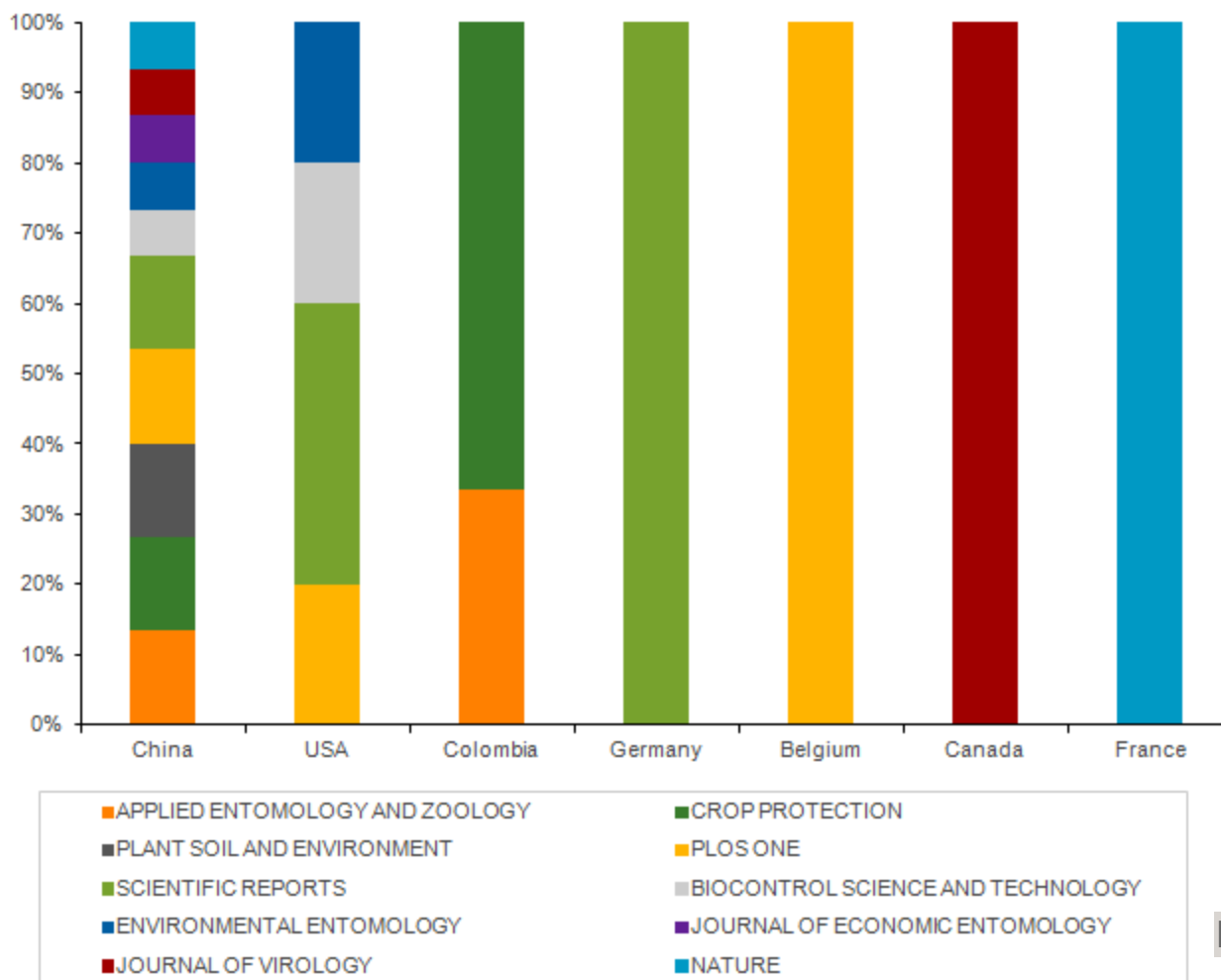
# I 科技论文发展概况

逐年主要发文期刊



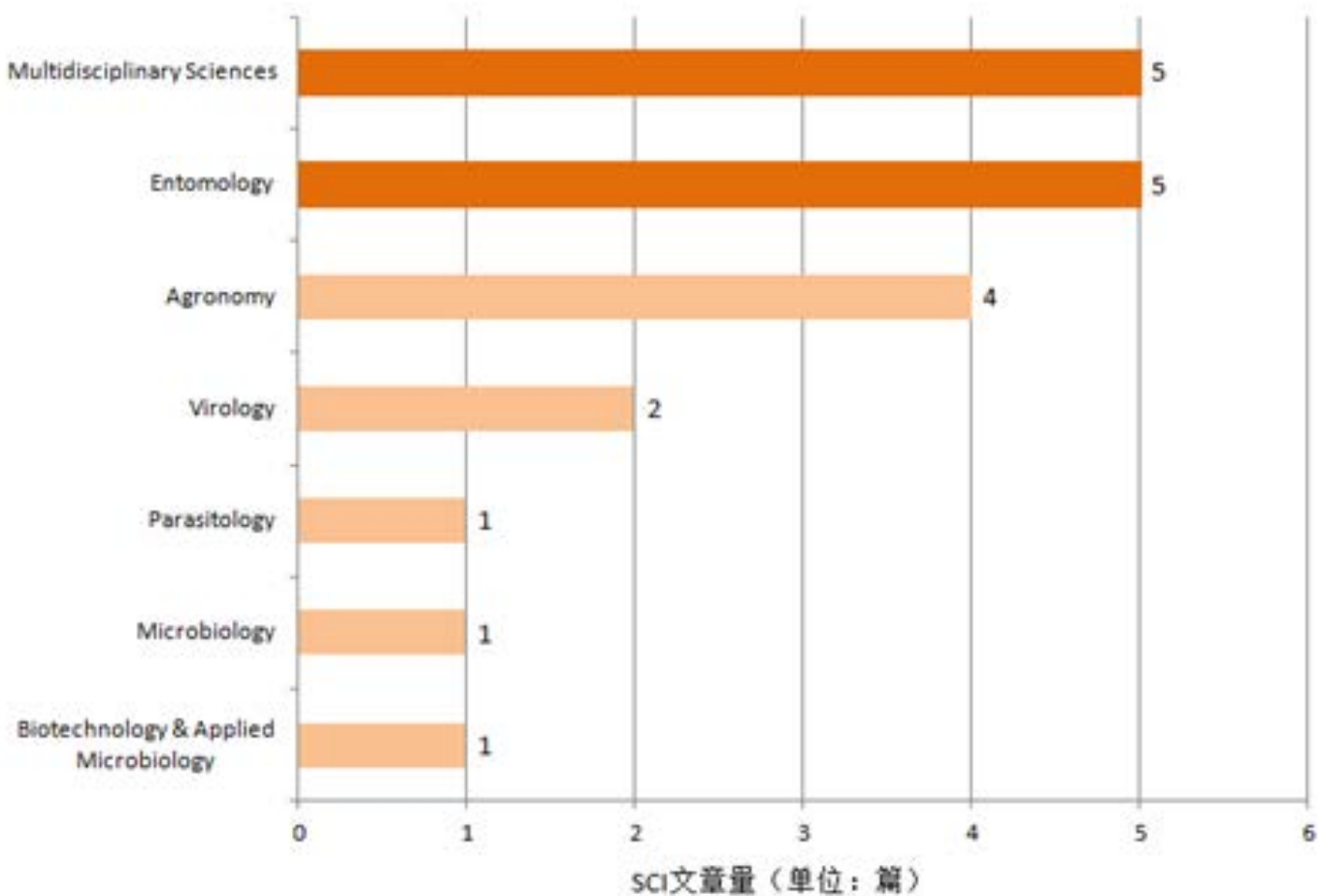
# I 科技论文发展概况

与各国地区合作的文章主要期刊分布



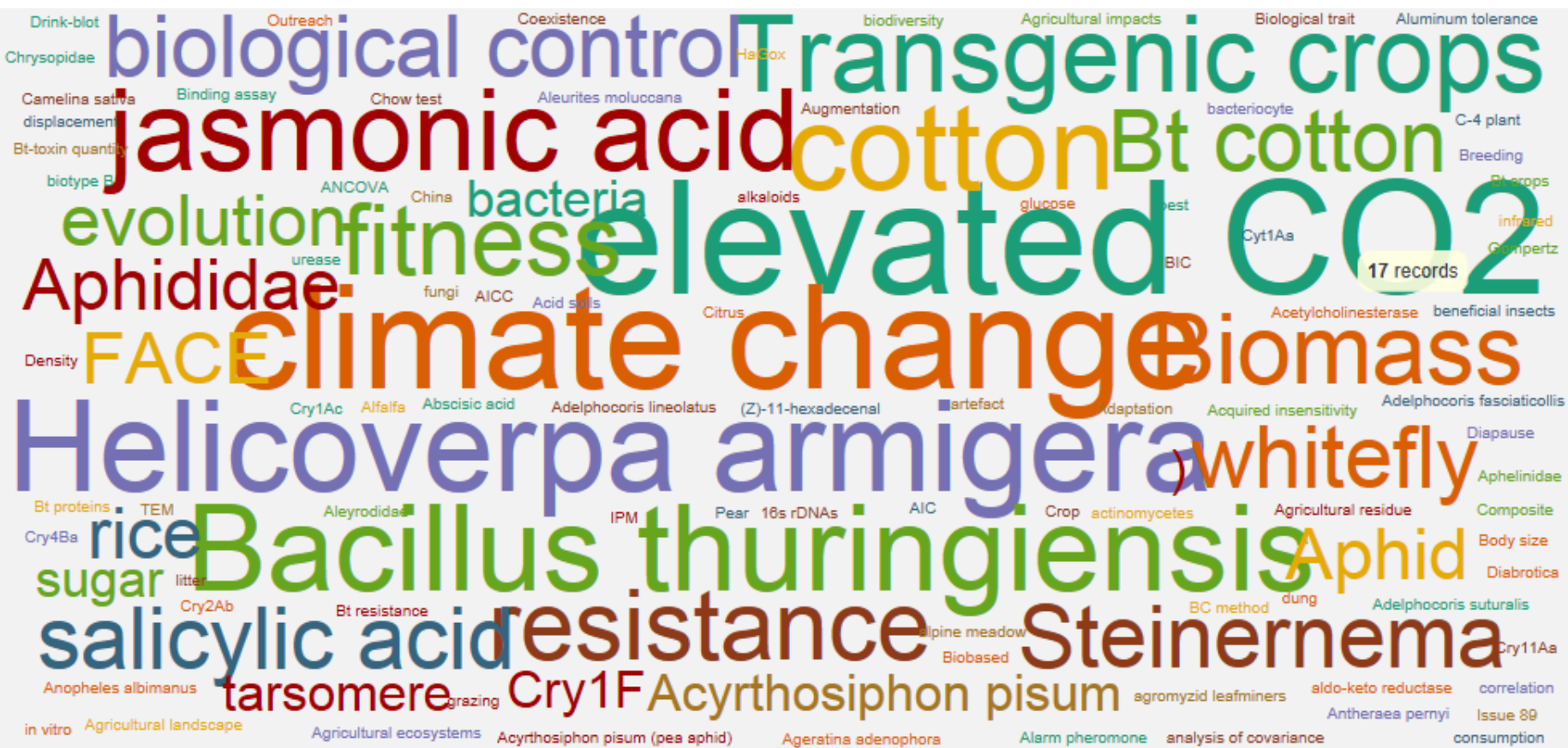
# I 科技论文发展概况

基于学科的研究重点 (Web of Science类别, 单位: 篇)



# I 科技论文发展概况

基于关键词的研究重点（利用作者关键词）



比如高频出现的词有二氧化碳浓度升高（elevated CO2，17条记录）  
气候变化（climate change，15条记录）等

# I 科技论文发展概况

## 最近3年新出现或消失的机构（尤其适合领域分析）

### Organizational Trends in Last 3 Years

Last 3 Years are: 2014 - 2012

Lists of the organizations in the dataset that have been the most active in the last three years (relative to the entire dataset in the last three years), only have activity in the last three years, or have no activity in the last three years.

#### Top Organizations in Last 3 Years

Chinese Acad Agr Sci [11]  
Chinese Acad Sci [2]  
Max Planck Inst Chem Ecol [2]  
Oregon State Univ [2]  
Univ Arizona [2]  
Natl Agrotech Extens & Serv Ctr [1]  
French Natl Inst Agr Res INRA [1]  
Royal Ontario Museum [1]  
Univ Ghent [1]  
Univ Lancaster [1]  
Wuhan Inst Bioengn [1]

#### Organizations First Published in Last 3 Years

Chinese Acad Sci [2]  
Max Planck Inst Chem Ecol [2]  
Oregon State Univ [2]  
Univ Arizona [2]  
French Natl Inst Agr Res INRA [1]  
Royal Ontario Museum [1]  
Univ Ghent [1]  
Univ Lancaster [1]  
Wuhan Inst Bioengn [1]

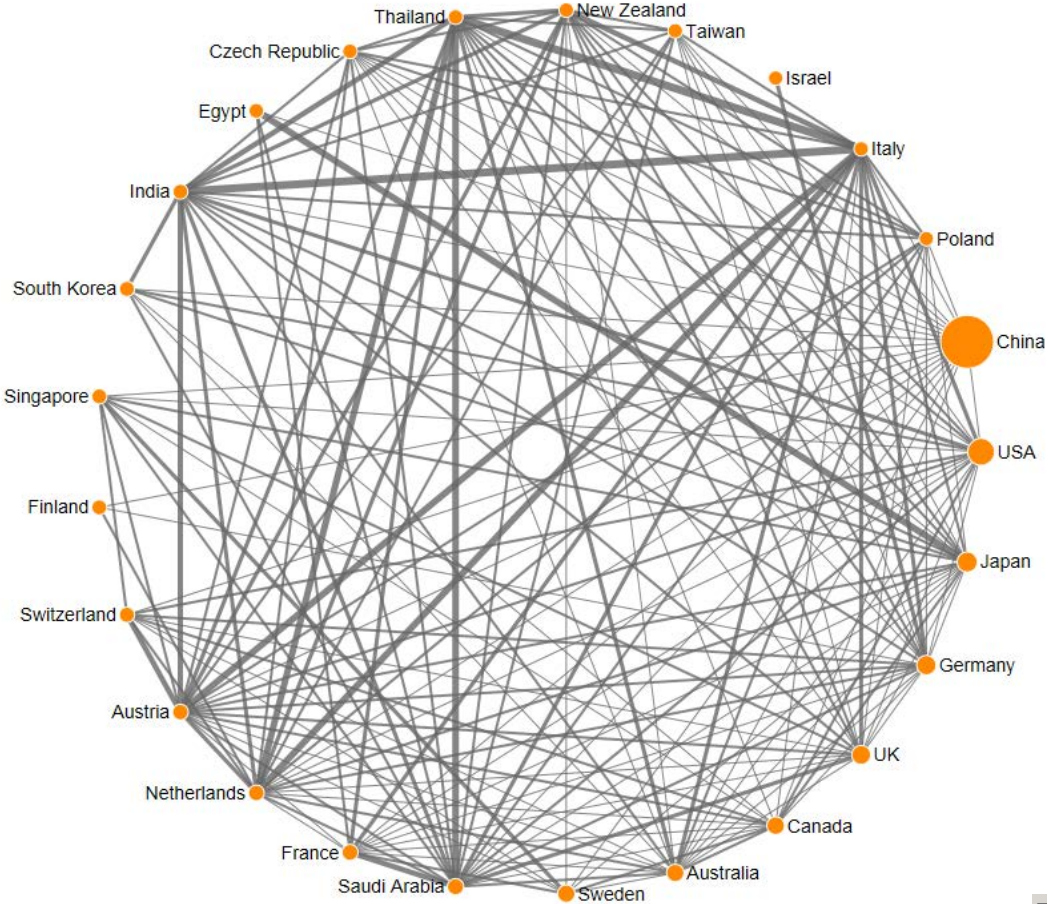
#### Organizations No Longer Published in Last 3 Years

Univ Bogota Jorge Tadeo Lozano [2]  
Henan Acad Agr Sci [1]  
Int Ctr Trop Agr CIAT [1]  
Univ Georgia [1]

# II 科研合作

## 基于国家层面的科研合作

合作较多的国家/地区（合作发文量 ≥ 10篇）

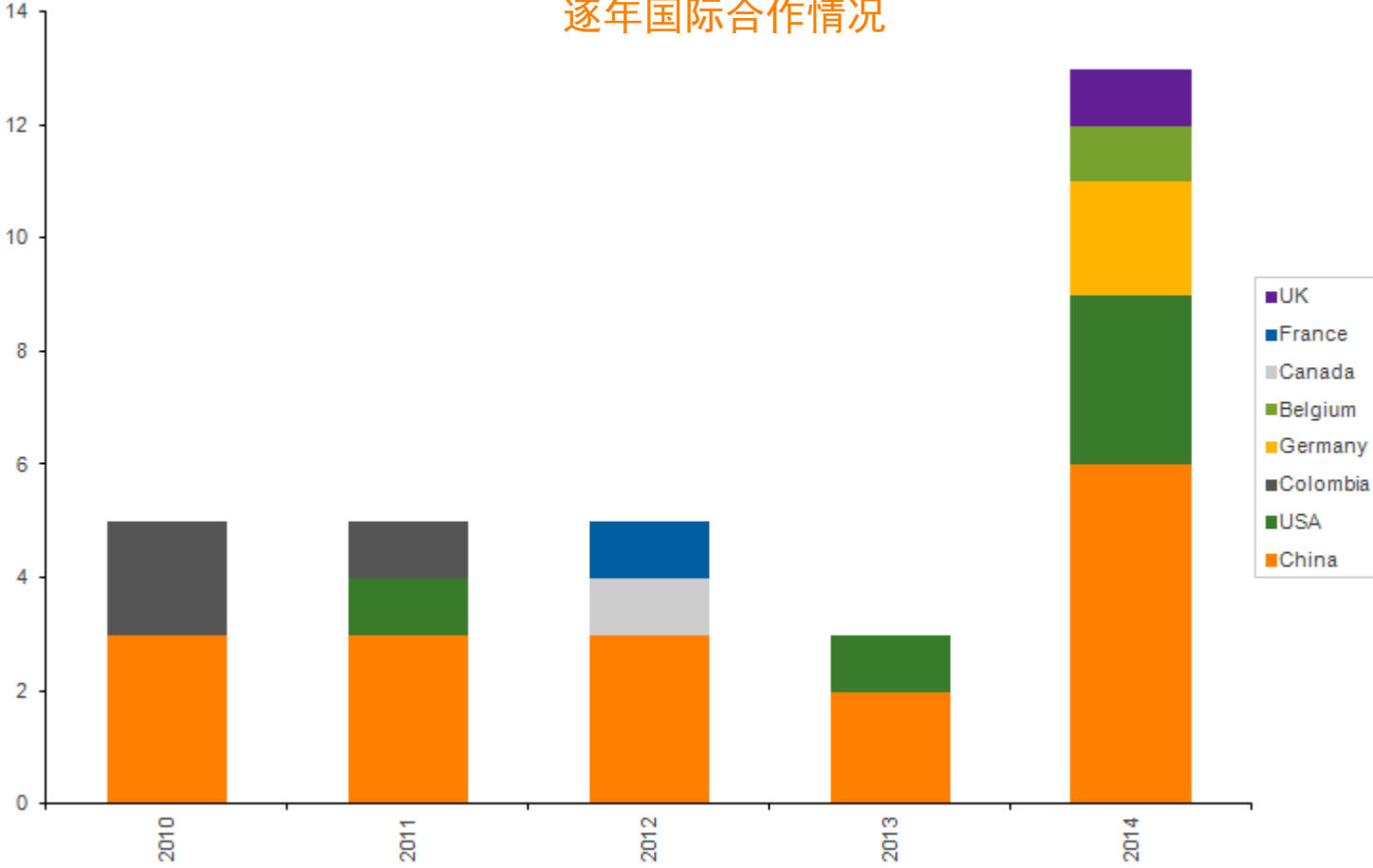




# II 科研合作

基于国家层面的科研合作

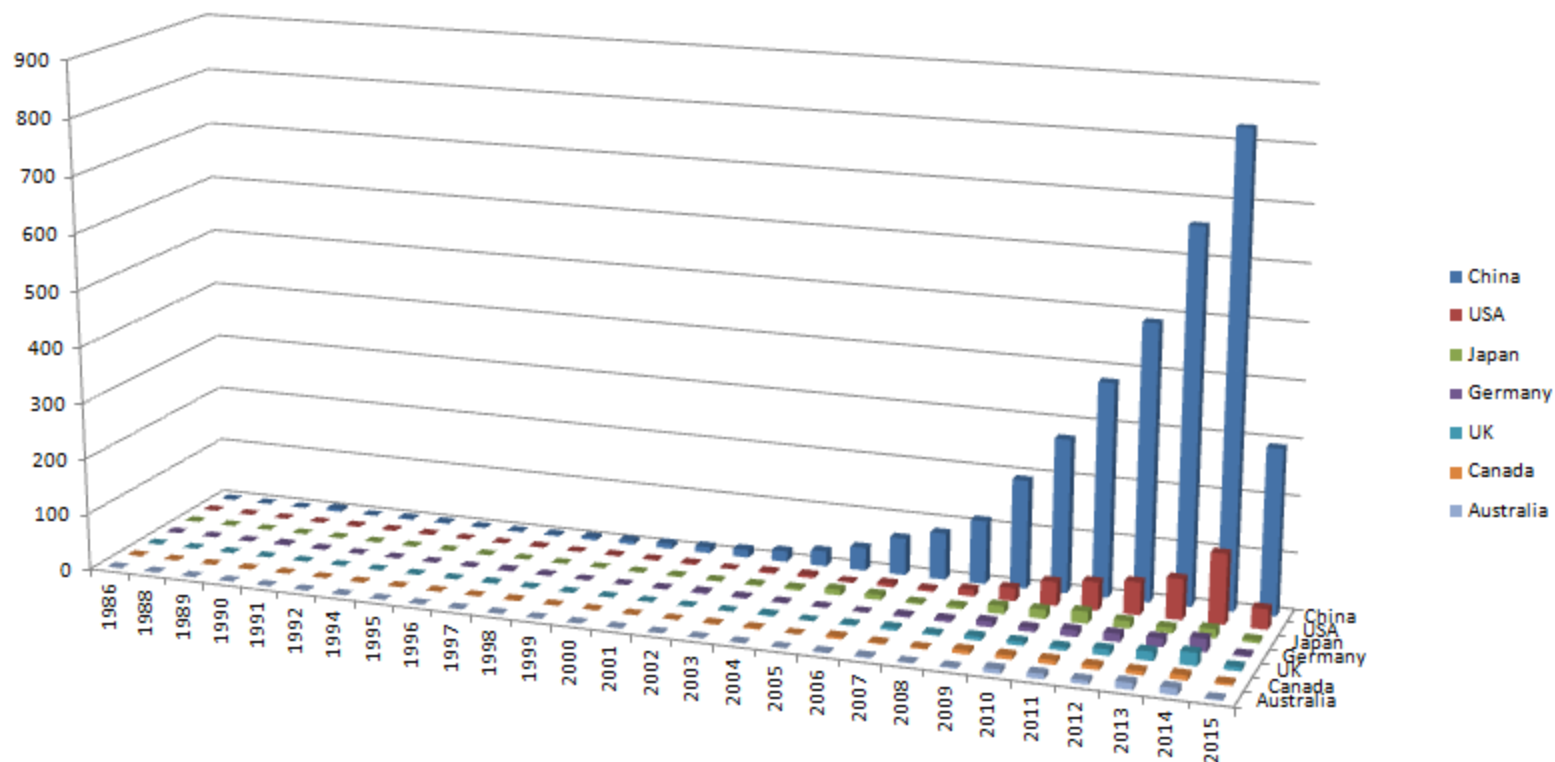
逐年国际合作情况



## II 科研合作

### 基于国家层面的科研合作

本机构或团队逐年国际合作情况（TOP7）



# II 科研合作

## 基于国家层面的科研合作

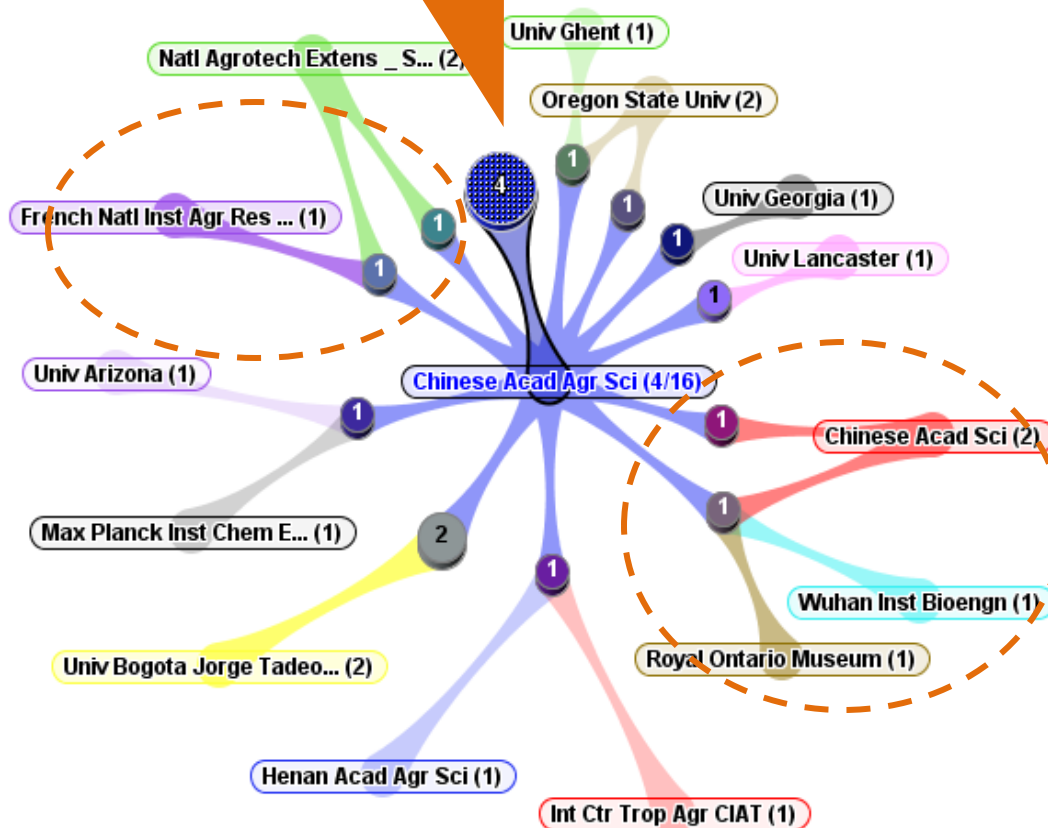
本团队国际合作逐年分布情况 (Top15高合作量的国家/地区)

Reset	Countries		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	# Records		2	2	1	1	3	2	1	1	5	4	8	12	13	20	22	29	71	96	11	12	14	20
Publication Year	# Records	▼ ▲		Show Values >= 1 and <= 202																				
		Cooccurrence # of Records		1990	1991	1992	1994	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
1	866	China	1		1	1	2	2	1	1	4	4	8	12	13	20	22	29	71	96	11	12	14	20
2	408	USA		1	1	1	1		1	1	2		3	5	1	6	4	12	23	42	50	58	72	12
3	125	Japan									1	1	2	4	10	9	3	5	14	16	21	13	10	16
4	103	Germany	2	1			2	1			1	3	2	1	1		3	5	10	7	12	14	15	23
5	83	UK					1							2		2	5	3	7	7	6	11	16	23
6	57	Canada						1					1	2		4	2	2	7	7	8	7	7	9
7	54	Australia														1	2	1	2	8	9	7	12	12
8	51	Sweden											1		1				5	8	9	8	9	10
9	26	Saudi Arabia																					14	12
10	24	France									1							3	2	1	3	2	5	7
11	23	Austria															2	2	2	3	1	4	4	5
12	22	Netherlands															1			3	1	3	6	8
13	19	Switzerland																	4	3	4	1	4	3
14	16	South Korea										1	1				2		1	1	3	2	2	3
15	15	Finland													1		1	1	1		1	3	4	3
16	15	Singapore																	1		4	1	5	4

## II 科研合作

### 基于机构层面的科研合作

XXX团队独立发表4篇文章



	各年度发文总量	3	3	3	2	5
合作文章总量	机构	2010	2011	2012	2013	2014
2	中国科学院			1	1	
2	全国农业技术推广服务中心（中国农业部直属事业单位）	1		1		
2	美国俄勒冈州立大学				1	1
2	哥伦比亚波哥大罗萨里奥大学	2				
1	法国国家农业研究院			1		
1	河南省农业科学院		1			
1	哥伦比亚国际热带农业中心		1			
1	德国马克斯·普朗克化学生态学研究所					1
1	皇家安大略博物馆			1		
1	（美国）亚利桑那大学					1
1	（美国）乔治亚大学		1			
1	（比利时）根特大学					1
1	（英国）兰卡斯特大学					1
1	武汉生物工程学院			1		



有4篇是某团队独立完成的科研成果，其余12篇文章都存在不同程度的机构合作。比如，与全国农业技术推广服务中心共合作2篇文章，其中有1篇还共同与法国国家农业研究院存在合作；与中科院共同合作发表2篇文章，其中有1篇还共同与武汉生物工程学院、皇家安大略博物馆存在合作

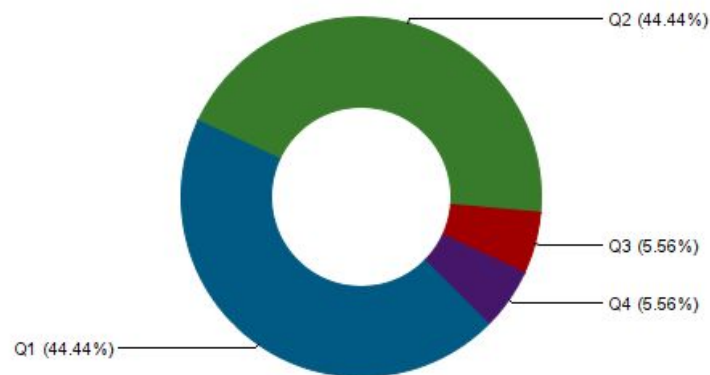
# III 学术质量分析

- ❖ 基于发文期刊
- ❖ 基于论文引用
- ❖ 基于引文数据

## 本团队SCI文章所在期刊、影响因子及其区间排名等信息

来源出版物	发文量	2014年 影响因子	期刊所属WOS学科	影响因子 所在区间
APPLIED ENTOMOLOGY AND ZOOLOGY	2	1.144	Entomology	Q2
CROP PROTECTION	2	1.493	Agronomy	Q2
PLANT SOIL AND ENVIRONMENT	2	1.226	Agronomy	Q2
PLOS ONE	2	3.234	Multidisciplinary Sciences	Q1
BIOCONTROL SCIENCE AND TECHNOLOGY	1	0.938	Biotechnology & Applied Microbiology Entomology	Q4 Q3
ENVIRONMENTAL ENTOMOLOGY	1	1.295	Entomology	Q2
JOURNAL OF ECONOMIC ENTOMOLOGY	1	1.506	Entomology	Q2
JOURNAL OF VIROLOGY	1	4.439	Virology	Q1
NATURE	1	41.456	Multidisciplinary Sciences	Q1
PLOS PATHOGENS	1	7.562	Microbiology Parasitology Virology	Q1 Q1 Q1
SCIENCE	1	33.611	Multidisciplinary Sciences	Q1
SCIENTIFIC REPORTS	1	5.578	Multidisciplinary Sciences	Q1

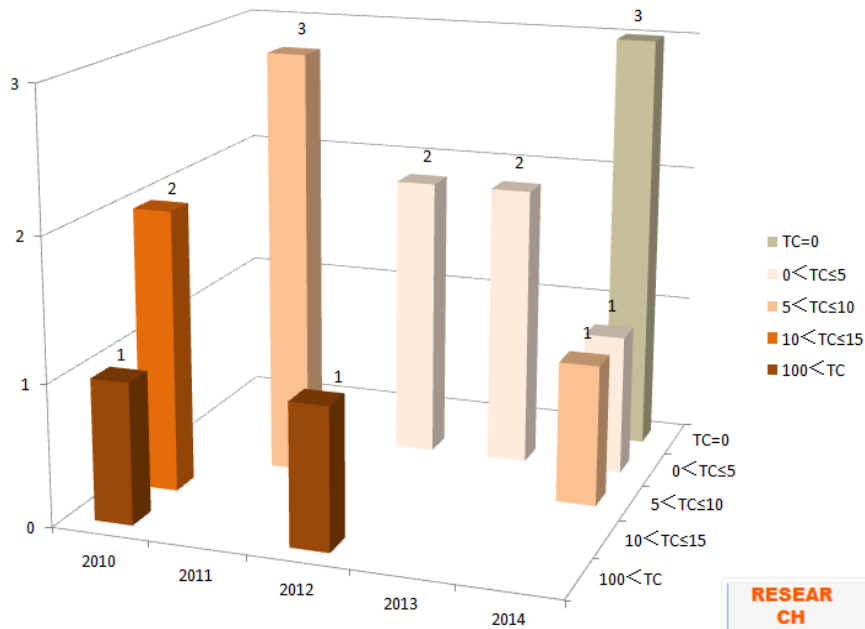
SCI文章所在期刊的区间统计



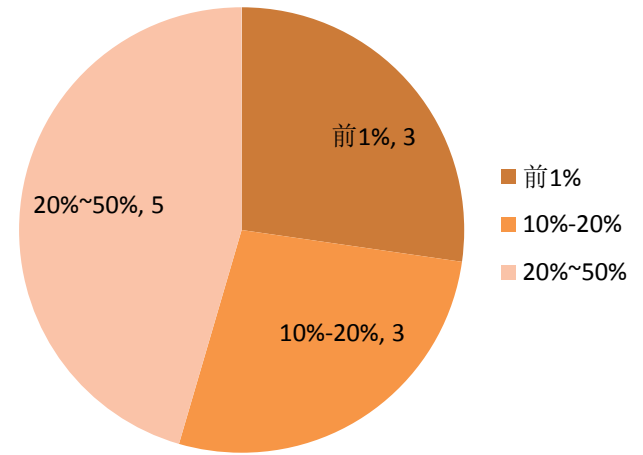
# III 学术质量分析

- ❖ 基于发文期刊
- ❖ 基于论文引用
- ❖ 基于引文数据

所发表的SCI文章的被引用情况



SCI文章所在期刊的区间统计



Report-Plot Matrix in Excel



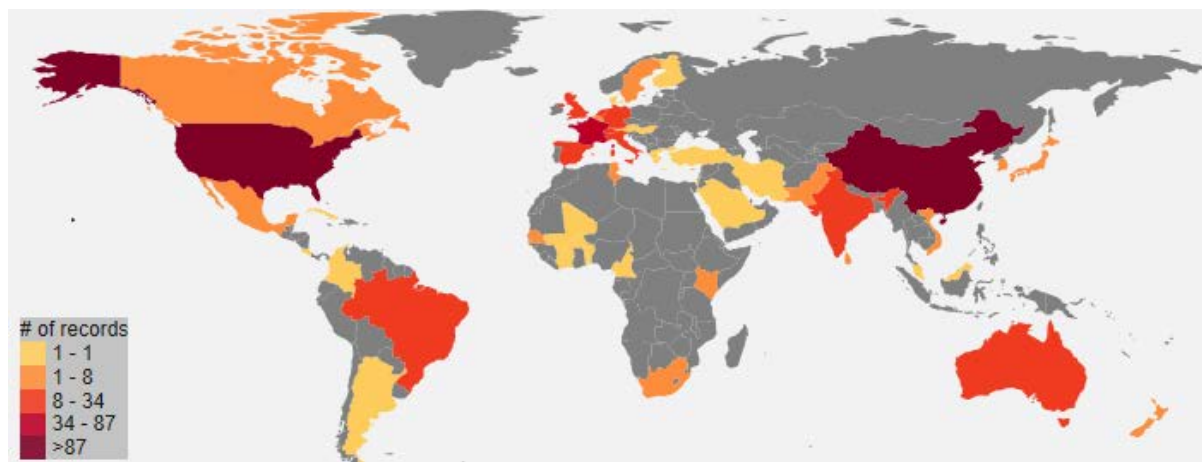
RESEARCH FIELD	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
CLINICAL MEDICINE	0.01%	2,414	1,585	1,615	1,373	1,439	1,097	926	784	445
	0.10%	742	698	581	517	472	368	272	223	139
	1.00%	235	206	185	162	140	116	90	66	44
	10.00%	63	57	50	44	39	33	26	20	13
	20.00%	38	35	30	27	24	20	16	12	8
	50.00%	14	13	11	10	9	8	6	5	3
COMPUTER SCIENCE										

数据来源: ESI, 数据采集时间: 20160405

# III 学术质量分析

- ❖ 基于发文期刊
- ❖ 基于论文引用
- ❖ 基于引文数据

关注本团队SCI科研成果的国家/地区（发文量≥2篇）

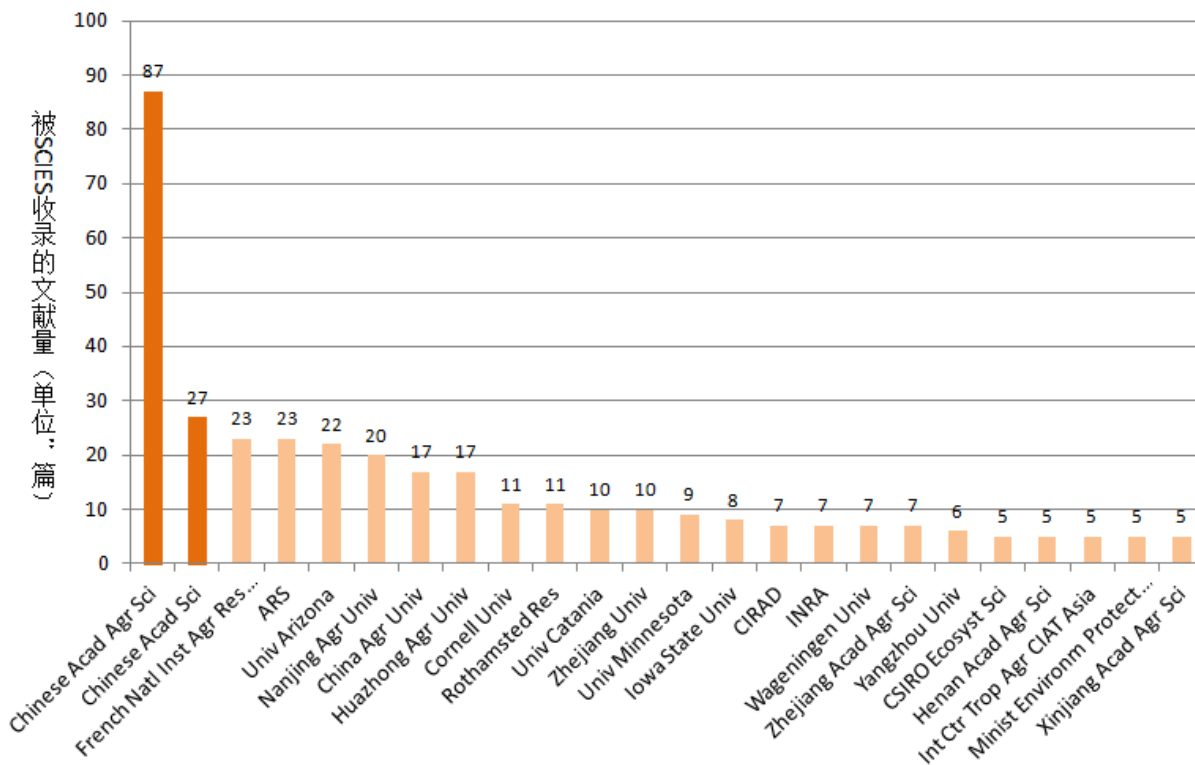


国家/地区	文章量	国家/地区	文章量	国家/地区	文章量	国家/地区	文章量
China	174	India	11	Vietnam	7	South Africa	3
USA	112	Germany	10	Mexico	5	South Korea	3
France	44	Netherlands	9	Japan	4	Senegal	2
UK	30	Spain	9	Sweden	4	Sri Lanka	2
Brazil	19	Switzerland	9	Kenya	3	Tunisia	2
Australia	18	Belgium	7	New Zealand	3		
Italy	16	Canada	7	Pakistan	3		

# III 学术质量分析

- ❖ 基于发文期刊
- ❖ 基于论文引用
- ❖ 基于引文数据

## 关注本团队SCI科研成果的机构（发文量≥5篇）



除中国农业科学院外，主要有中国科学院（27篇）、法国国家农业研究院（23篇）、美国农业部农业研究局（23篇）、（美国）亚利桑那大学（22篇）、南京农业大学（20篇）、中国农业大学（17篇）、华中农业大学（17篇）、（美国）康奈尔大学（11篇）、英国洛桑研究所（11篇）、（意大利）卡塔尼亚大学（10篇）、浙江大学（10篇）、（美国）明尼苏达大学（9篇）、（美国）爱荷华州立大学（8篇）等

Report-Plot List in Excel



# III 学术质量分析

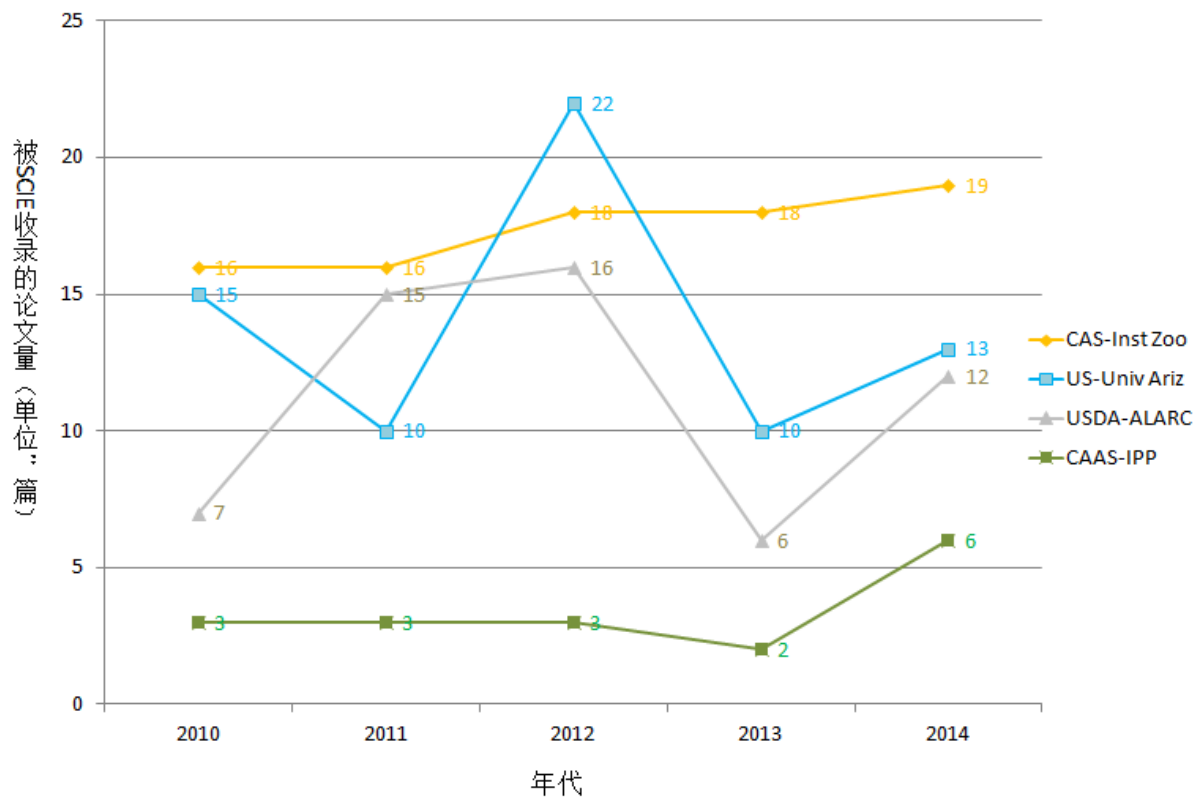
- ❖ 基于发文期刊
- ❖ 基于论文引用
- ❖ 基于引文数据

## 施引文献中ESI高被引论文的国家及机构分布

	文章总量	3	3	2	2	1	1	1	1
文章总量	施引文献中ESI高被引论文的署名单位	法国	美国	比利时	西班牙	中国	意大利	墨西哥	英国
2	French Natl Inst Agr Res INRA	2		1	1		1	1	
2	Univ Arizona	1	2			1			
2	Univ Ghent	1	1	2	1		1		
1	Chinese Acad Agr Sci		1			1			
1	Cornell Univ		1	1					
1	Ctr Cooperat Int Rech Agron Dev	1	1						
1	InVivo AgroSolut	1						1	
1	Iowa State Univ		1	1					
1	Kansas State Univ		1	1					
1	Nanjing Agr Univ		1			1			
1	Passeig Lluís Co 23				1				1
1	Univ Catania	1		1	1		1		
1	Univ Guadalajara	1						1	
1	Univ Lleida				1				1
1	Univ Nebraska		1	1					
1	Univ Politecn Madrid	1		1	1		1		
1	Univ Warwick				1				1
1	USDA ARS		1	1					

## IV 对标分析

四支团队的逐年发文情况（2010-2014）



## IV 对标分析

### 四支团队科研合作的国家分布（2010-2014）

CAAS-IPP (共10篇)		CAS-Inst Zoo (共37篇)		US-Univ Ariz (共36篇)		USDA-ALARC (共20篇)	
USA	5	USA	24	China	17	China	8
Colombia	3	South Africa	6	Canada	9	Canada	6
Germany	2	Netherlands	4	France	5	Switzerland	5
Canada	1	Denmark	3	Germany	3	Japan	2
Belgium	1	Italy	3	Israel	3	Belgium	2
France	1	Japan	2	Mexico	3	Denmark	1
UK	1	France	1	UK	2	Italy	1
		India	1	South Africa	1	France	1
		Sweden	1	Denmark	1	India	1
				Japan	1	Germany	1
				Netherlands	1	Australia	1
				India	1	Spain	1
				Australia	1	Nepal	1
				Spain	1	Philippines	1
				Cameroon	1	South Korea	1
				Chile	1		
				New Zealand	1		
				Reunion	1		
				Turkey	1		

## IV 对标分析

四支团队所发表文章的被引频次区间分布

	发文量	87	67	54	16
发文量	四支团队	CAS-Inst Zoo	US-Univ Ariz	USDA- ALARC	CAAS- IPP
22	TC=0	8	8	3	3
129	0<TC≤10	65	28	30	10
40	10<TC≤20	11	19	8	2
13	20<TC≤30	2	6	7	
3	30<TC≤40	1	1	1	
4	40<TC≤50		2	2	
2	50<TC≤60		2		
3	60<TC≤70		1	2	
1	80<TC≤90		1		
1	90<TC≤100		1		
5	100< TC≤150		1	3	1
1	150< TC≤200				1

# IV 对标分析

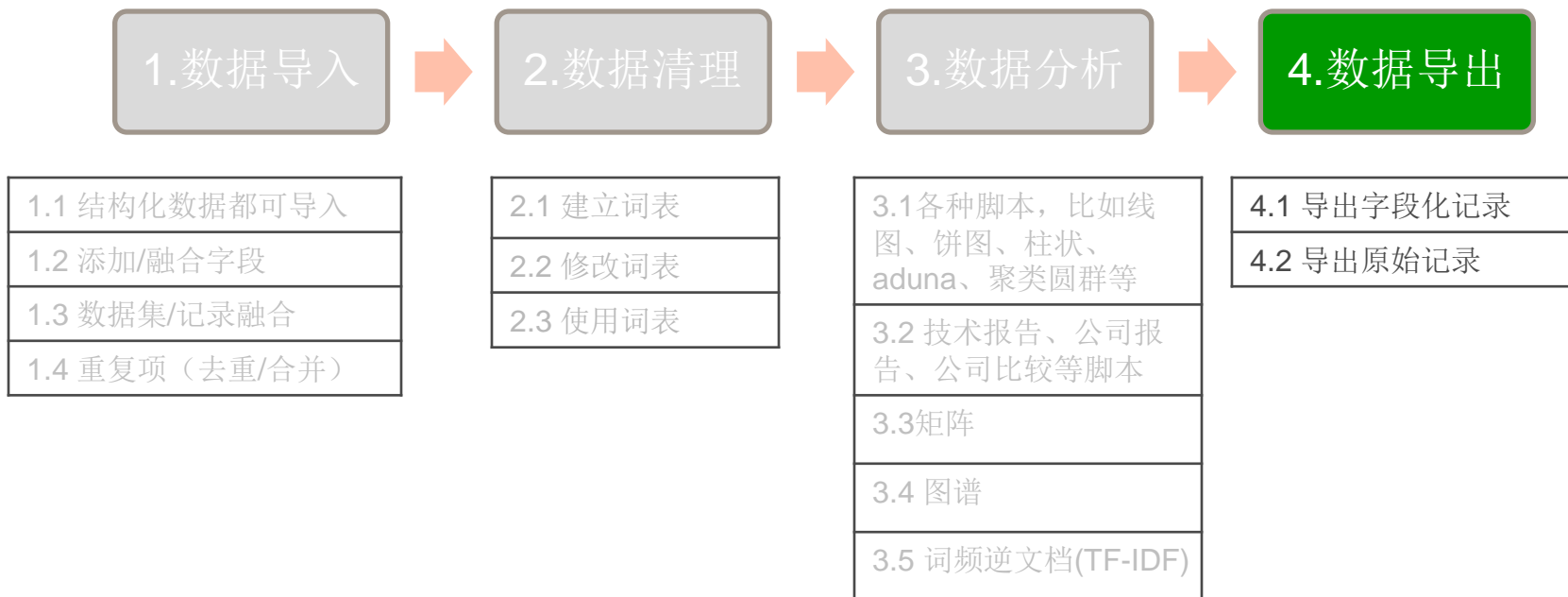
## 四支团队基于学科的科研重点分析

Unique Article Identifier	# Records	# Records	# Records			
			CAS-Inst Zoo	US-Univ Ariz	USDA-ALARC	CAAS-IPP
1	72	Entomology	39	16	12	5
2	45	Multidisciplinary Sciences	11	28	6	6
3	26	Agronomy	3		19	4
4	21	Ecology	10	2	9	
5	20	Zoology	8	12		
6	15	Biochemistry & Molecular Biology	11	2	2	
7	15	Plant Sciences	12		3	
8	13	Environmental Sciences	5		8	
9	10	Biology	9		1	
10	9	Biotechnology & Applied Microbiology	2	3	3	1
11	8	Physiology	7	1		
12	6	Agricultural Engineering			6	
13	6	Microbiology		5		1
14	5	Biodiversity Conservation	1		4	
15	5	Meteorology & Atmospheric Sciences			5	
16	4	Agriculture, Multidisciplinary	1		3	
17	4	Biochemical Research Methods	1		3	
18	4	Evolutionary Biology		3	1	
19	4	Soil Science	4			
20	3	Forestry			3	
21	3	Parasitology		2		1
22	3	Virology		1		2
23	2	Behavioral Sciences	1	1		
24	1	Chemistry, Applied				1
25	1	Food Science & Technology				1
26	1	Imaging Science & Photographic Technolo		1		
27	1	Medicine, General & Internal		1		
28	1	Neurosciences	1			
29	1	Remote Sensing		1		
30	1	Toxicology	1			

## 四支团队基于主题词的科研重点分析

发文总量 (单位: 篇)	主题词	CAS-Inst Zoo	US-Univ Ariz	USDA-ALARC	CAAS-IPP
117	昆虫/害虫	55	42	14	10
79	作物	36	17	24	2
41	抗性/抗性治理	7	29	6	3
38	效果/影响	16	10	11	3
28	CO <sub>2</sub> 浓度升高	26		2	
26	Bt蛋白	2	16	6	2
24	国家/地区	5	9	5	5
22	模型/建模	9	6	9	
18	适应/适应性	8	9	1	
17	气候变化/变暖	2		15	
9	反应	7		2	
8	细菌		7		1
7	天敌	4		3	
7	捕食者	3		4	
7	温度	5		1	1
7	病毒	4	1		2
7	茉莉酸	7			
6	充裕、丰富 (Abundance)	6			
6	红外线保温/加热			6	

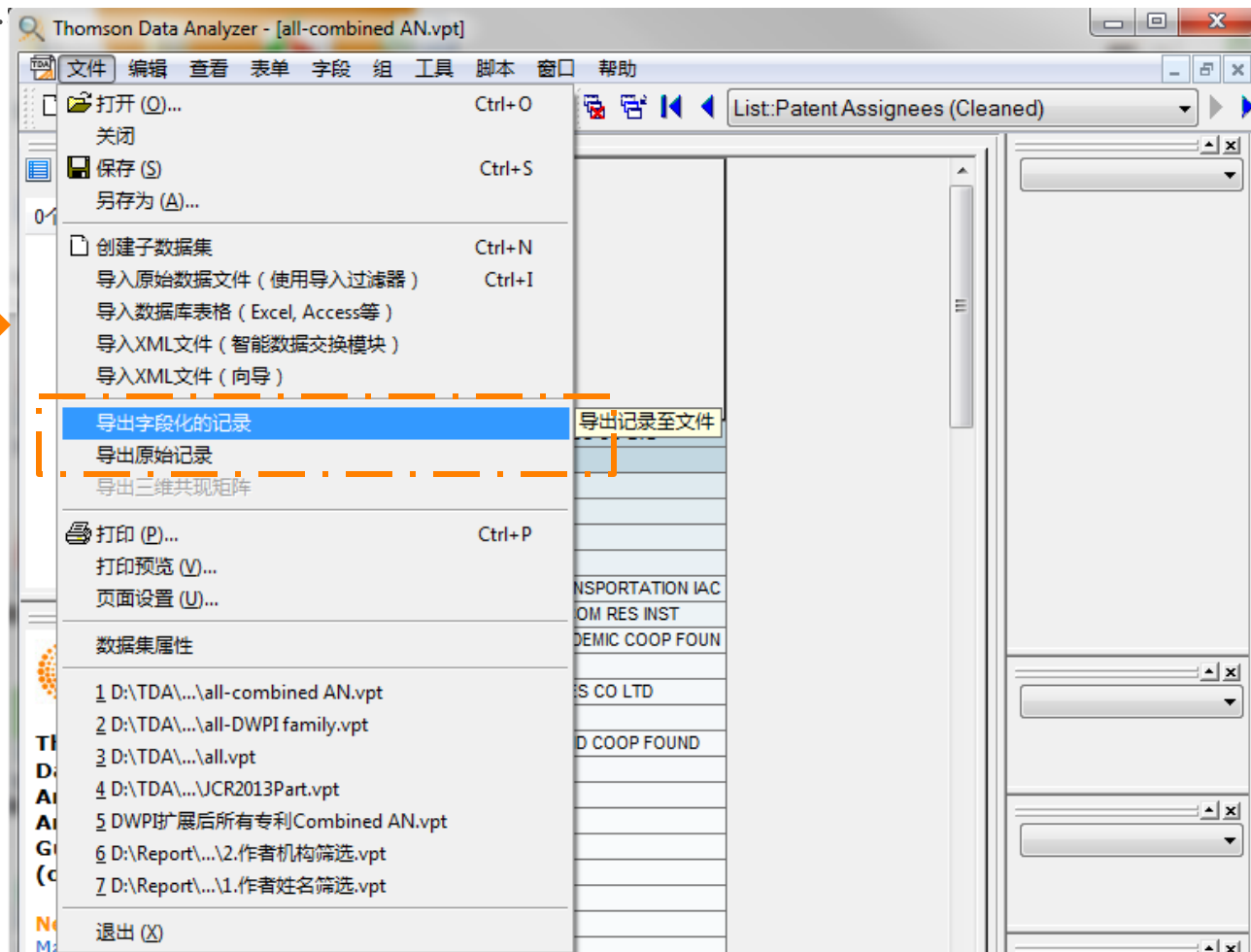
# TDA简要流程概览



## 4. 数据导出

4.1 导出字段化记录

4.2 导出原始记录



# TDA简要流程概览

## 1.数据导入

- |                |
|----------------|
| 1.1 结构化数据都可导入  |
| 1.2 添加/融合字段    |
| 1.3 数据集/记录融合   |
| 1.4 重复项（去重/合并） |

## 2.数据清理

- |          |
|----------|
| 2.1 建立词表 |
| 2.2 修改词表 |
| 2.3 使用词表 |

## 3.数据分析

- |                                 |
|---------------------------------|
| 3.1 各种脚本，比如线图、饼图、柱状、aduna、聚类圆群等 |
| 3.2 技术报告、公司报告、公司比较等脚本           |
| 3.3 矩阵                          |
| 3.4 图谱                          |
| 3.5 词频逆文档(TF-IDF)               |

## 4.数据导出

- |             |
|-------------|
| 4.1 导出字段化记录 |
| 4.2 导出原始记录  |





# Thomson Data Analyzer

---

1. TDA主要功能
2. TDA分析实例
3. TDA常见问题 Q&A



# TDA能解决哪些问题？

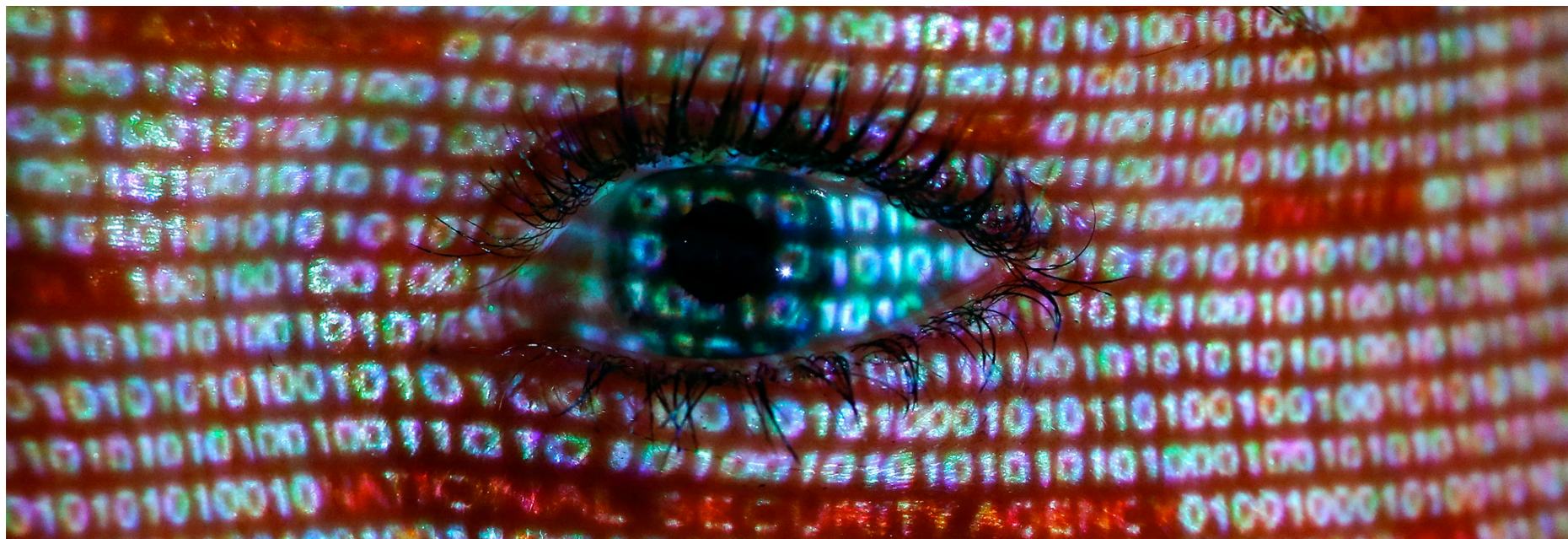
高效解决问题	
1	TDA能分析什么数据？ CNKI/EI可以导入TDA吗？
2	在导出数据到TDA时，落掉某个字段，后续还可以追加进来吗？
3	能把本机构作为第一署名单位或通讯作者地址的文章挑出来吗？ 怎么知道高产出的第一作者是谁？（TDA中有第一作者全称，比缩写参考性更高）
4	能挑出本校所有Q1区间的论文吗？
5	能挑出本机构属于ESI临床医学的所有文章吗？
6	刚整理好2015年度SCI发文信息，能整合进入历史机构知识库中吗？
7	能去重吗？
8	能清理什么字段？（机构、人、文本等）



## TDA能解决哪些问题？

高效解决问题	
9	清理之后的结构能保存为词表再次利用吗？之前的词表能修改吗？
10	手边的机构变体表/专业词汇能转置为TDA识别的词表吗？
11	TDA中的数据可以Excel表形式导出带走
12	做分析报告，可在TDA中直接用数据出Excel表中的常规图，比如树图、饼图、线图、泡泡图、环图等
13	地图形式呈现国家地区分布
14	有多少文章是目前尚未被引用的，有多少的引用次数是 > 50次的？
15	能看本机构和谁合作的多吗？
16	在研究本机构或某领域的研究重点时，除了关注学科外，还想看看关键词。关键词从哪找？从什么角度分析？





REUTERS/Pawel Kopczynski

## 联系我们！

---

技术支持: [ts.support.china@thomsonreuters.com](mailto:ts.support.china@thomsonreuters.com)

Tel: 4008 822 031(工作时间：周一至周五, 9:00—17:00)

Fax: 010-82862088

北京市海淀区科学院南路2号 融科资讯中心C座北楼610室  
汤森路透知识产权与科技集团



THOMSON REUTERS