



REUTERS/Pawel Kopczynski

# Thomson Data Analyzer (TDA) 的高效利用

---

张丹丹

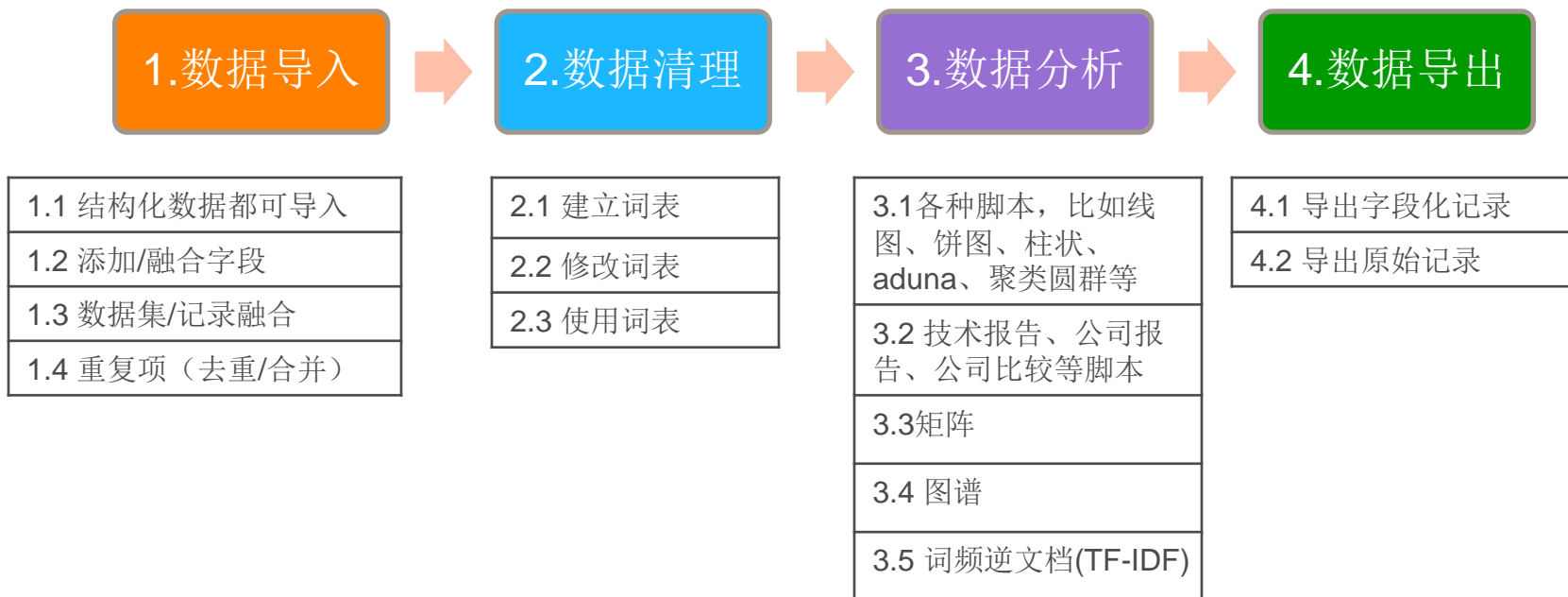
产品与解决方案部

汤森路透



THOMSON REUTERS

# TDA的工作机理



# TDA简要流程概览

## 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简要流程概览

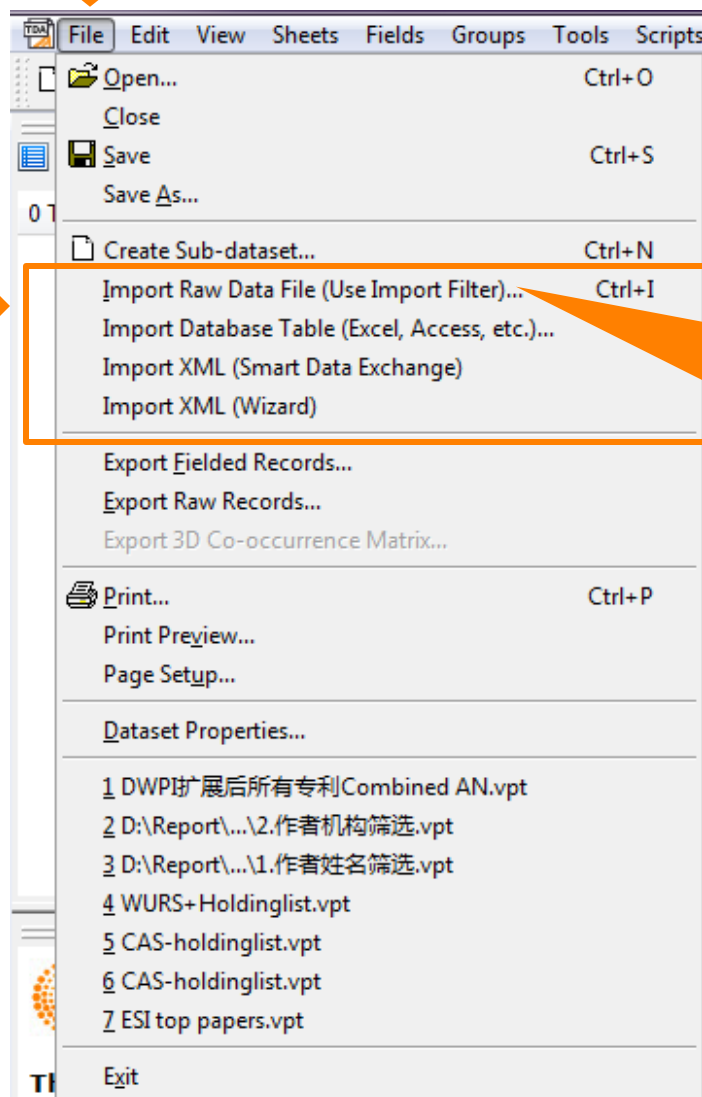
## 1.数据导入

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简要流程概览

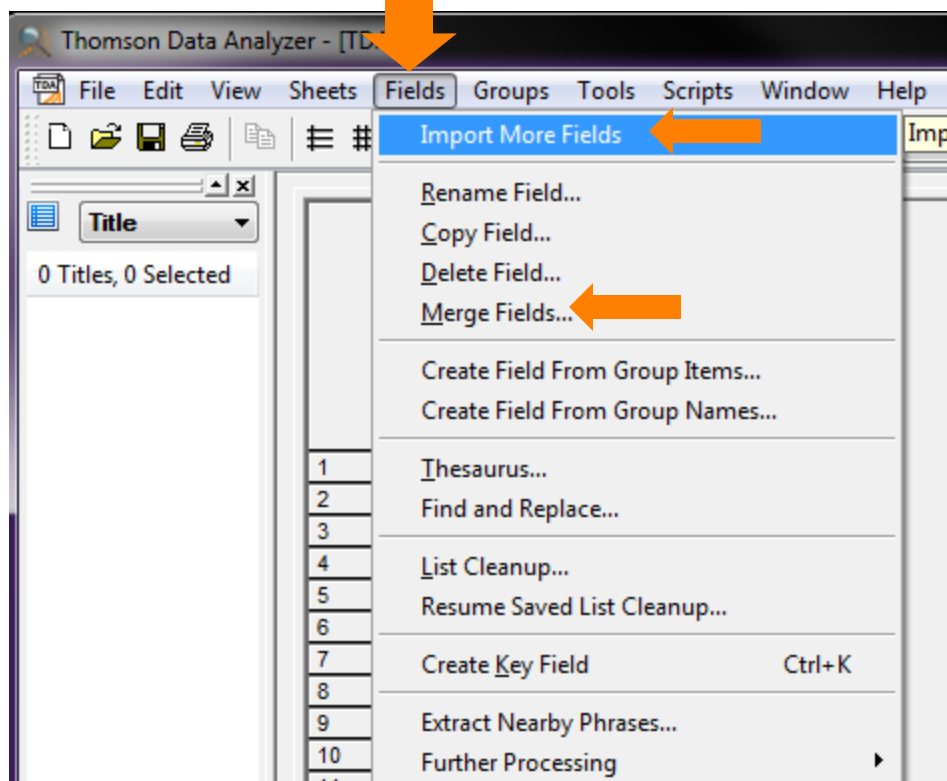
## 1.数据导入

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

Field	Number of Items
(filters)	>=500
1st author's	281
Author Affiliations (1st)	42
Author Affiliations (1st) (Full)	169
Author Affiliations (Authors - Organization)	653
Author	
Organization	
Author Affiliations (City and Country)	134
Author Affiliations (Full)	484
Author Affiliations (Organization and City and Country)	234
Author Affiliations (Organization Only)	218
Authors	1,168
Authors (1st)	259
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 Thesaurus
- Utility - Mark Unpublished

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 Liu, Yi
2	2	2	Beijing Forestry Univ, Beijing Key Lab Lignocellulos Chem, Beijing 100083, Peoples R China Dai, Lin
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简要流程概览

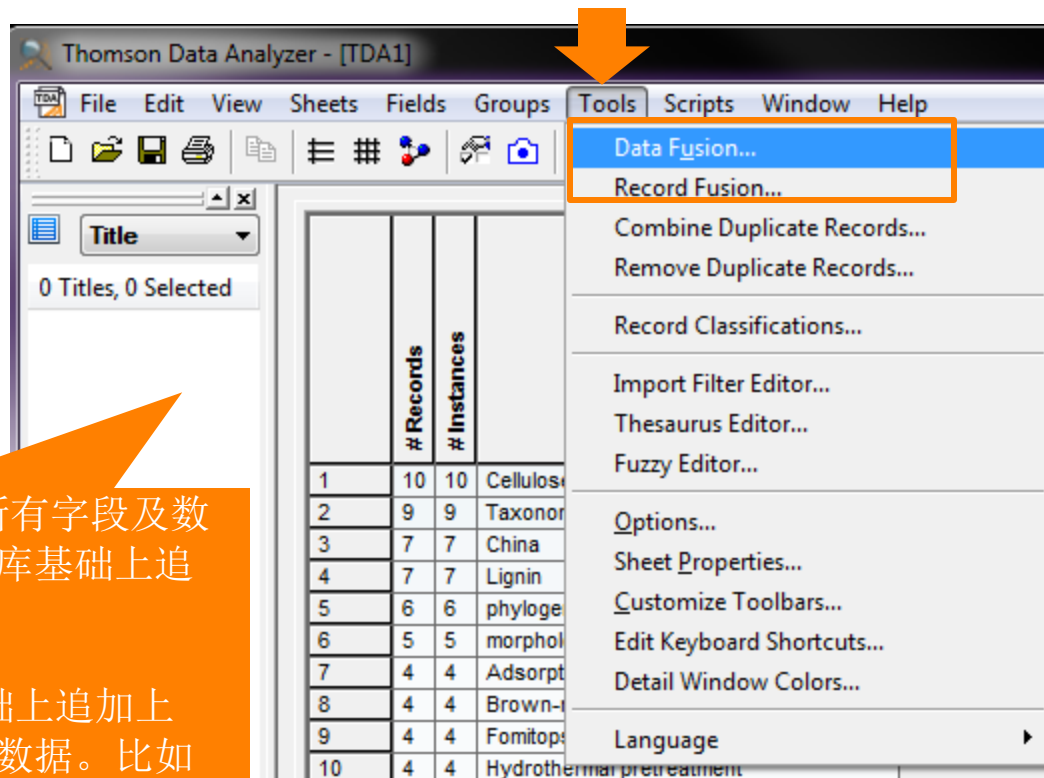
## 1.数据导入

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简要流程概览

## 1.数据导入

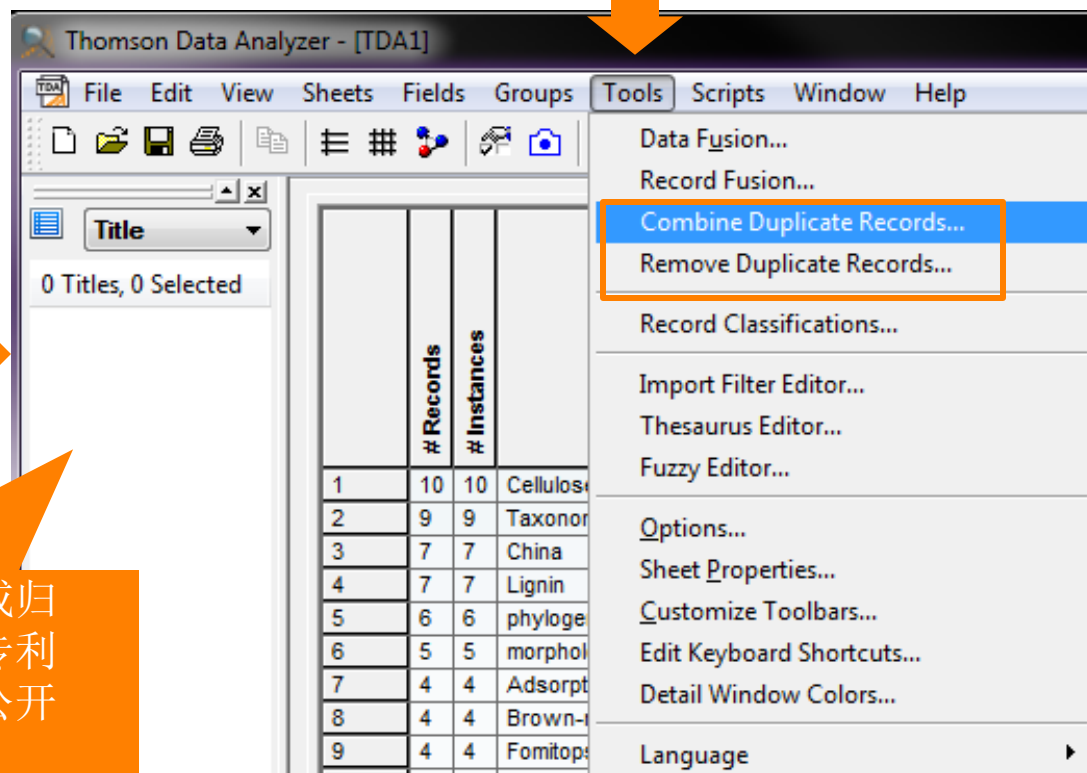
1.1 结构化数据均可导入

1.2 添加/融合字段

1.3 数据集/记录融合

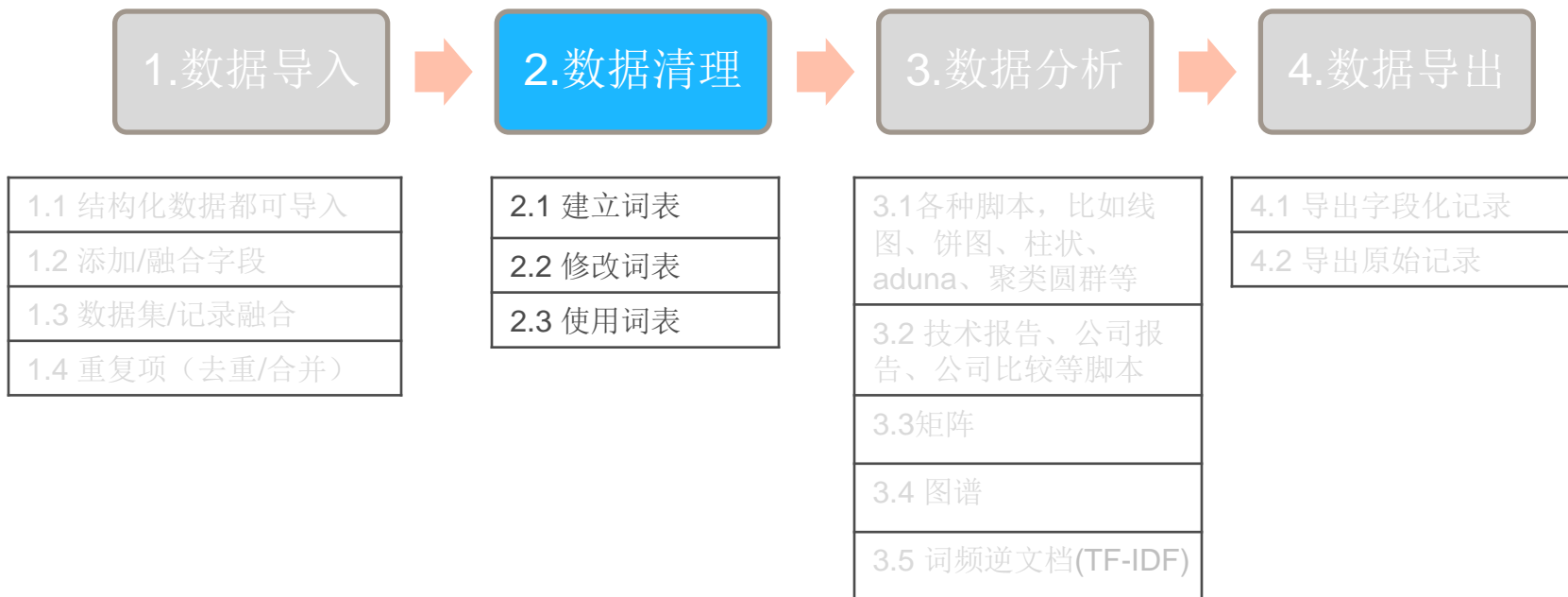
1.4 重复项（去重/合并）

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简要流程概览



# TDA简要流程概览

---

## 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)。

# 动画研究领域主要国家/地区的研究主要方向

Reset		Countries (Group Names)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25		
		# Records	31	37	13	11	11	11	32	20	7	7	66	14	8	52	10	13	93	8	26	23	10	15	10	35	20		
Keywords (author's) + Keywords Plus (1)	# Records	<div><div>▼▲</div><div>Show Values &gt;= 1 and &lt;= 118</div><div>Cooccurrence # of Records</div><div>▼▲</div></div>	animation	virtual reality	simulation	3D Technology	model	COMPUTER GRAPHICS	deformation	fluids	galaxies: evolution	DIGITAL SKY SURVEY	Japan animation & manga	GPU	VISUAL SIMULATION	motion	Computer games	optimization	multitieracies	SMOKE	algorithms	motion capture	Water	collision detection	real-time	reality	knowledge		
		1	1031	USA	10	11	44	34	39	41	15	5	4	5	11	5	2	19	34	3	36	1	11	6	4	5	2	13	8
		2	381	UK	36	35	9	12	13	8	2	1	4	5	9	1		7	20	1	3		6	3	2			1	4
		3	174	CANADA	20	24	5	2	9	9	3	1	3	4	4	2		3	5		15	1				1	2	2	3
		4	172	GERMANY	18	28	8	13	10	4	3	3	4	4	1	1		3	6		2			1		2	1	1	
		5	160	AUSTRALIA	11	13	5	2	7	2			3	3	8	1	2		5		18								2
		6	152	JAPAN	19	12	2	6	5	8	1	2	4	5	10	2	1	4	2	3	2	1		2					1
		7	144	CHINA	33	22	15	14	14	8	6	6	6	6	5	5	5	5	5	4	4	4	4	4	4	4	4	4	3



# 动画研究领域主要国家/地区的研究主要方向

USA	UK	CANADA	GERMANY	AUSTRALIA	JAPAN	CHINA
虚拟现实	动画	虚拟现实	虚拟现实	多元化读写能力	动画	动画
动画	虚拟现实	动画	动画	虚拟现实	虚拟现实	虚拟现实
模拟	电脑游戏	多元化读写能力	3D技术	动画	日本动画	模拟
计算机图形学	观众	模型	模型	教育	“恐怖谷”	3D技术
模型	性能/表现	教育	增强现实	读写能力	计算机图形学	模型
多元化读写能力	模型	计算机图形学	“恐怖谷”	日本动画	日本	计算机图形学
3D技术	环境	读写能力	模拟	模型	系统	变形
电脑游戏	3D技术	多峰性	系统	教育学	3D技术	流体
性能/表现	感知	观众	虚拟环境	媒体	感知	数字化
环境	“恐怖谷”	性能/表现	面部	电脑游戏	模型	日本动画

# TDA简要流程概览

---



## 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简要流程概览

---

## 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）

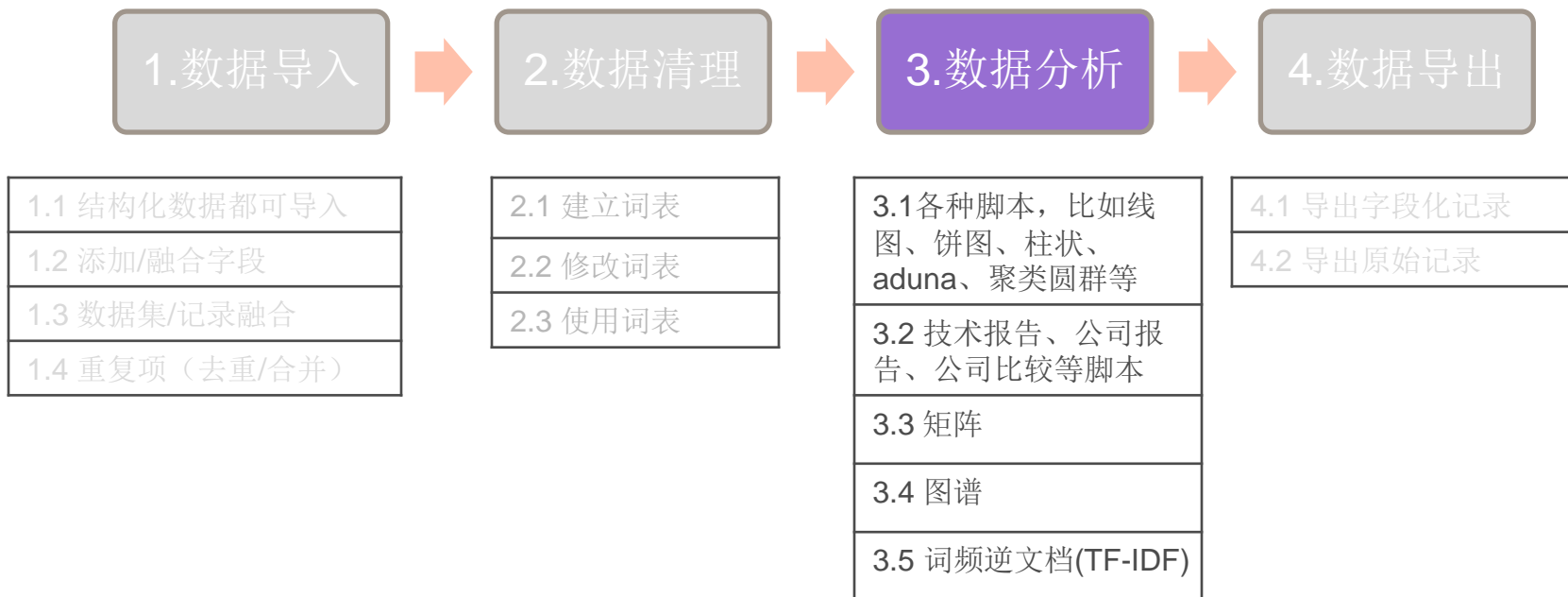
需要注意：

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。



# TDA简要流程概览



# TDA简要流程概览

## 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)

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简要流程概览

---

## 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)



### 词频逆文档 TF-IDF

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。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。



# TDA简要流程概览

## 3.数据分析

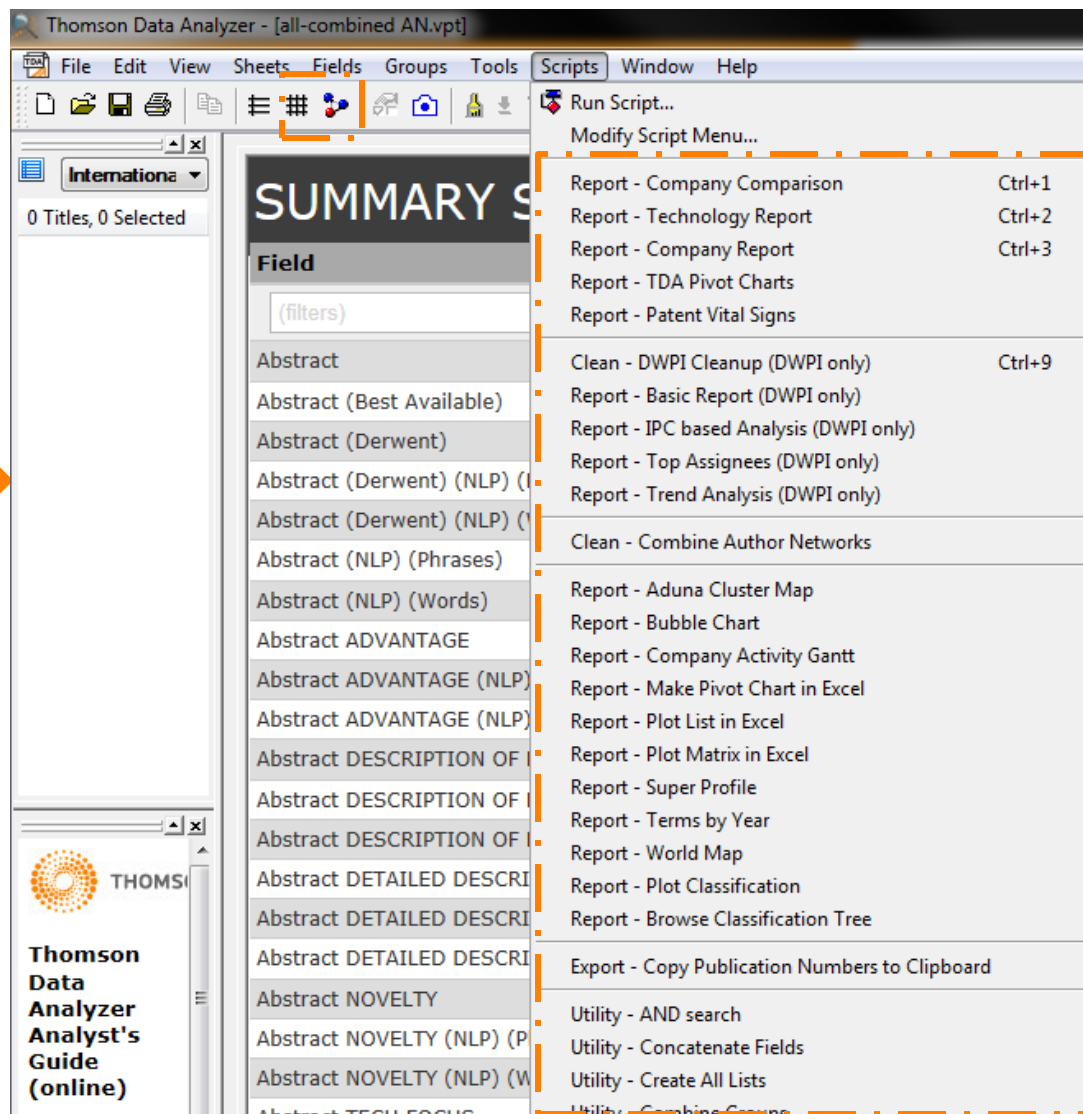
3.1 各种脚本，比如线图、饼图、柱状、aduna、聚类圆群等

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



# Technology Report:

TDA Tech Report.xls

	A	B	C	D	E	F
1						
2						
3		Summary Information				
4						
5		Year Chart				
6						
7		Organization Chart				
8						
9		Country Chart				
10						
11		Technology Chart				
12						
13		Technology v Year Chart				
14						
15		Technology v Country Chart				
16						
17		Country v Year Chart				
18						
19		Organization Holdings Chart				
20						
21		Organization Profile				
22						
23		Technology Profile				
24						
25		People Profile				
26						
27		Country Profile				
28						
29		Time Profile				
30						
31						
32						
33						
34						
35						
36						
37						
38						

**THOMSON REUTERS**

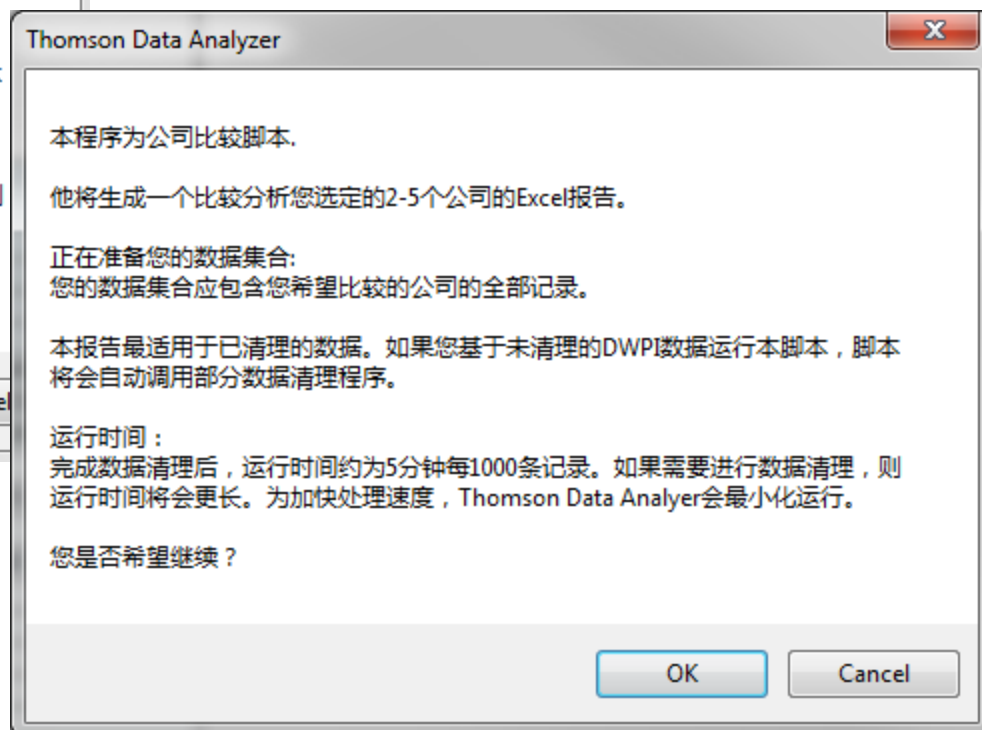
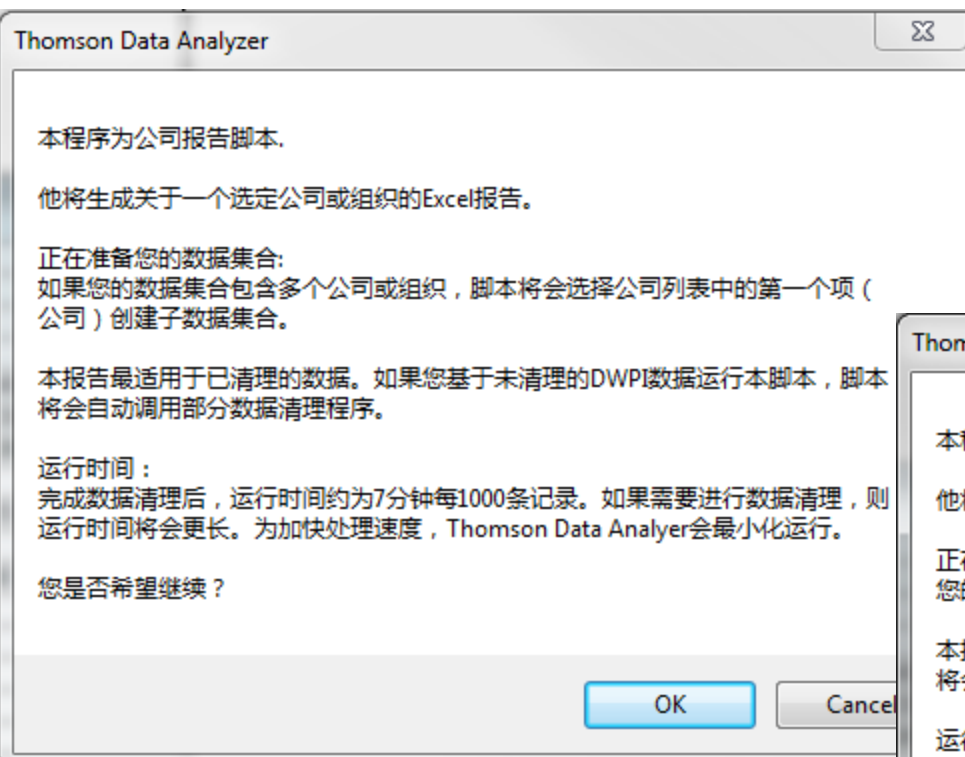
**THOMSON DATA ANALYZER**  
**Technology Report**

**Total Number of Records:** 139  
**Years Range From:** 2010 - 2015  
**Peak Year:** 2013 [78 Records]  
**Date of report creation:** 11/30/2015  
**Technology:** Manual Codes

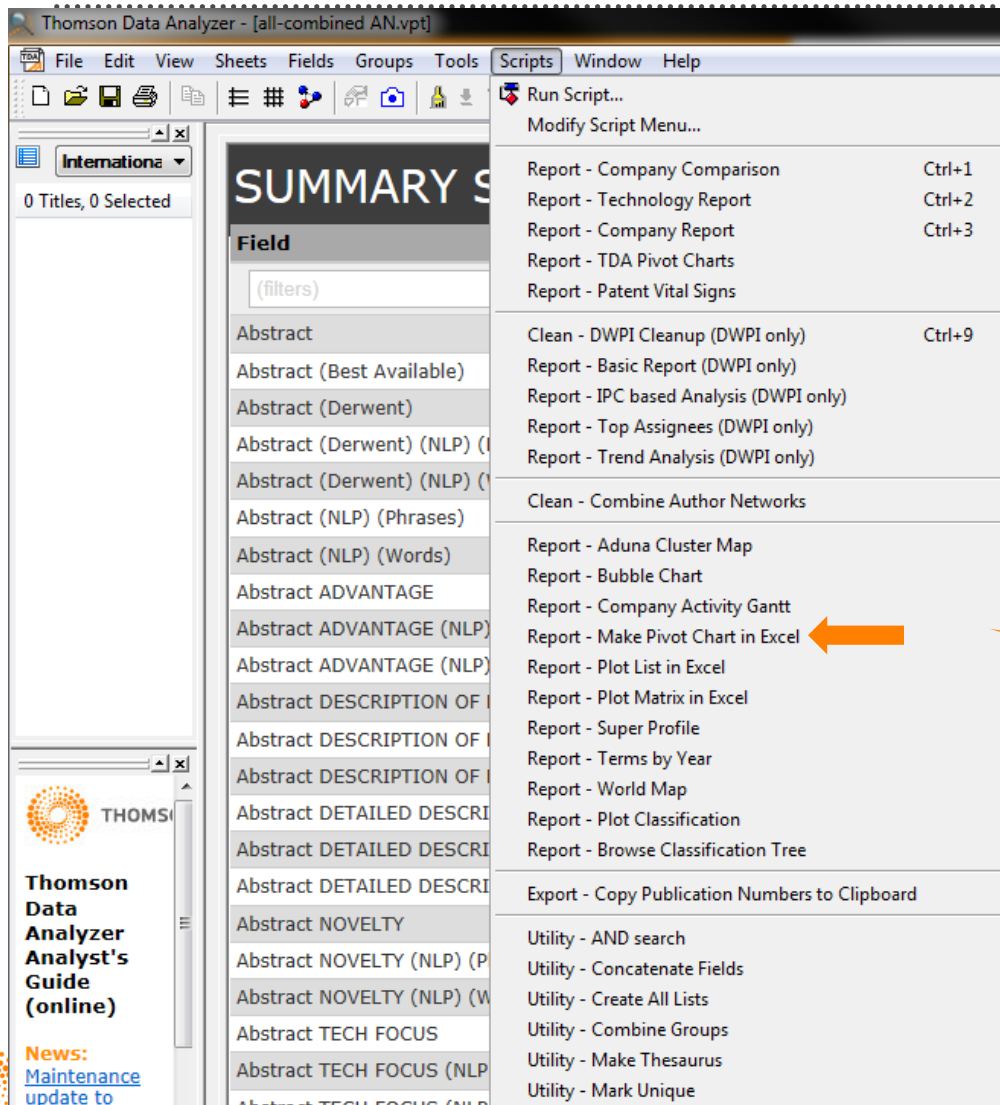
**Introduction** | Summary Info | Year Chart | Organization Chart | Country Chart | Technology Chart | Country v Year Chart | Technology v Year Chart | Technology v Country Chart



# Company Report / Company Comparison



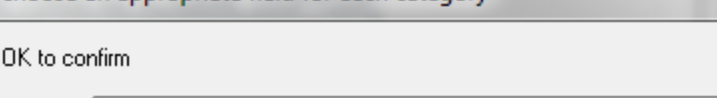
# Make Pivot Chart in Excel



比如查看各国家市场每年中的主要专利权人

# Make Pivot Chart in Excel

比如查看各国家市场每年中的主要专利权人



Please choose an appropriate field for each category

Click OK to confirm

Row Field: Application Countries

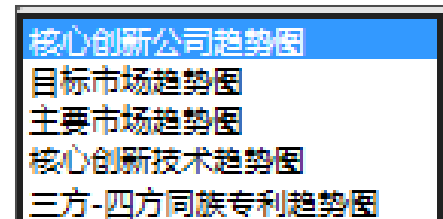
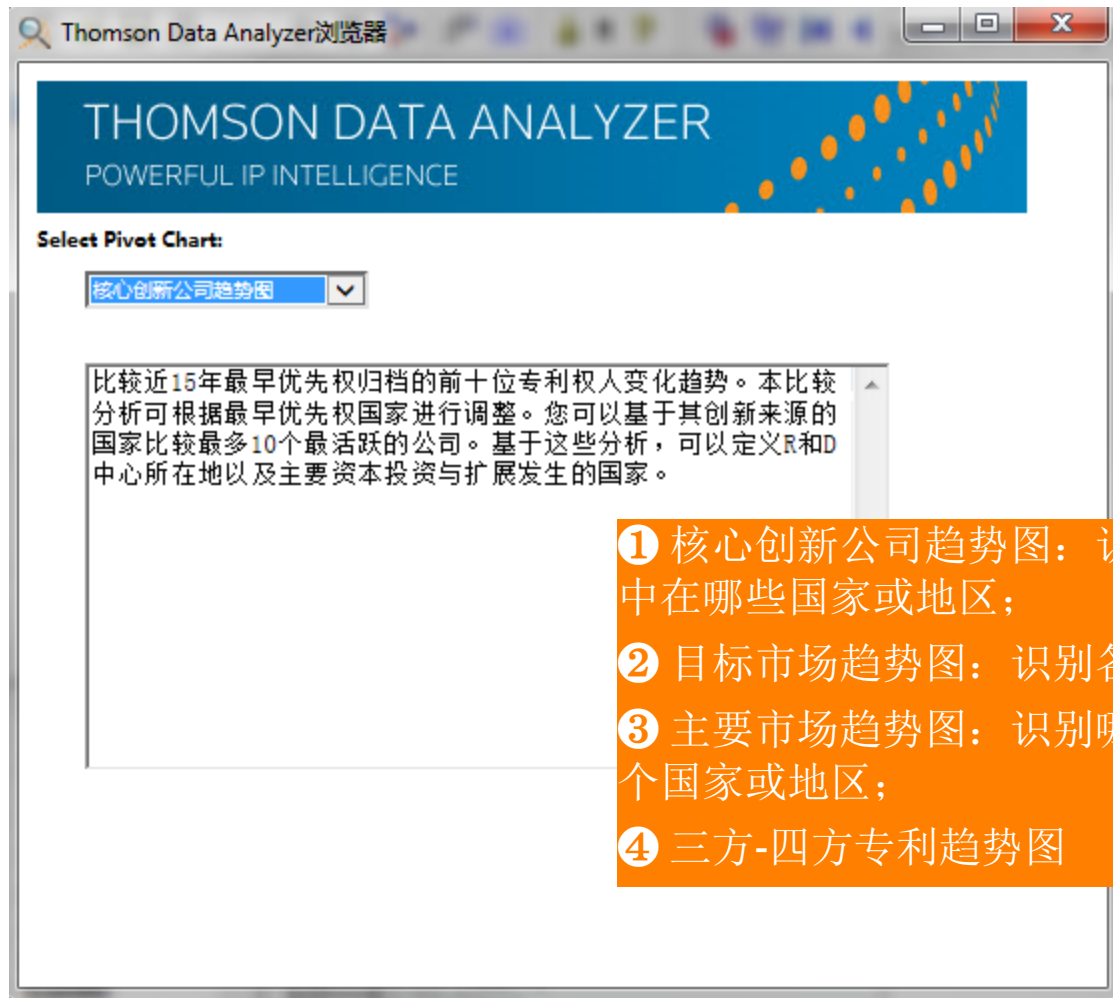
Column Field: Publication Year

Pivot Field: Patent Assignees (Cleaned)

Restore Defaults OK Cancel

Application Countries		Publication Year	Patent Assignees (Cleaned)	Num Records															
					Patent Assignees (Cleaned) (All)														
CN	2014	AHN J Y	1																
CN	2014	ALCATEL LUCENT	2		Sum of Num Records														
CN	2014	BEIJING INST TEC	1		Column Labels														
CN	2014	BROADCOM COR	1		Row Labels	AU	CA	CN	DE	EP	IN	JP	KR	TW	US	WO	Grand Total		
CN	2014	ELECTRONICS & T	1		2010			1									1		
CN	2014	NAM J Y	1		2012							1				4	5		
CN	2014	NEC LAB AMERIC	1		2013			7					10		11	12	40		
CN	2014	SAMSUNG ELECTI	2		2014			1	39	1	14		1	16	3	22	17	114	
CN	2014	SHANGHAI LANG	1		2015			1		8		5	4	7	6	2	39	35	107
CN	2014	UNIV BEIJING PO	2		Grand Total			1	1	55	1	19	4	9	32	5	72	68	267
CN	2014	UNIV BEIJING TEC	1																
CN	2014	UNIV BEIJING UN	2																
CN	2014	UNIV CHONGQIN	1																

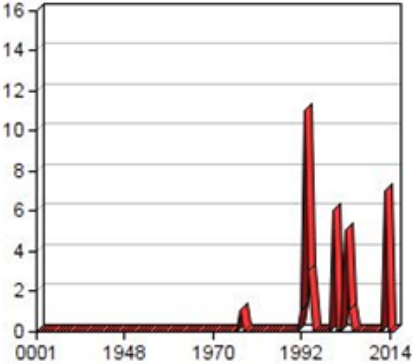
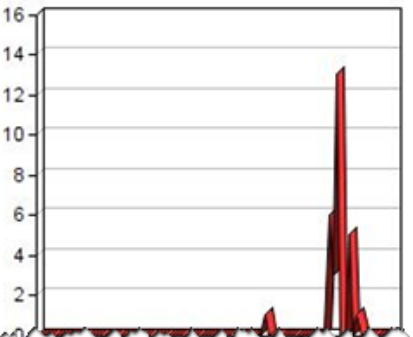
# Pivot Charts



- ① 核心创新公司趋势图：识别各公司的研发中心及投资集中在哪些国家或地区；
- ② 目标市场趋势图：识别各公司的主要生产和销售区域；
- ③ 主要市场趋势图：识别哪些公司主要定位在中日美欧四个国家或地区；
- ④ 三方-四方专利趋势图



# Super Profile

A	B	C	D												
Patent Assignees (Best Available) TOP10	Priority Years (earliest) Top 5 Items	Priority Years (earliest) Top Terms (minimum 5 record)	Application Years Year Line Chart												
FERRO CORP[35]	1992 [12]; 2012 [7]; 2002 [6]; 2000 [6]; 1993 [3]	1992 [12]; 2012 [7]; 2002 [6]; 2000 [6]	 <table border="1"><caption>Application Years for FERRO CORP</caption><thead><tr><th>Year</th><th>Count</th></tr></thead><tbody><tr><td>1992</td><td>11</td></tr><tr><td>2000</td><td>6</td></tr><tr><td>2002</td><td>6</td></tr><tr><td>2012</td><td>7</td></tr><tr><td>2014</td><td>7</td></tr></tbody></table>	Year	Count	1992	11	2000	6	2002	6	2012	7	2014	7
Year	Count														
1992	11														
2000	6														
2002	6														
2012	7														
2014	7														
NGK SPARK PLUG CO LTD[29]	2000 [16]; 1998 [7]; 2003 [4]; 1983 [1]; 1999 [1]	2000 [16]; 1998 [7]	 <table border="1"><caption>Application Years for NGK SPARK PLUG CO LTD</caption><thead><tr><th>Year</th><th>Count</th></tr></thead><tbody><tr><td>1998</td><td>7</td></tr><tr><td>2000</td><td>16</td></tr></tbody></table>	Year	Count	1998	7	2000	16						
Year	Count														
1998	7														
2000	16														



# Patent vital signs

[illegible]

# Export-Copy Publication Numbers to clipboard

Thomson Data Analyzer - [all-combined AN.vpt]

File Edit View Sheets Fields Groups Tools Scripts Window Help

Application

17 Titles, 0 Sel...

CN  
CN  
EP  
EP  
JP  
KR  
KR  
TW  
TW  
US

	# Records	# Instances	Inventors (Cleaned)	Derwent	Potential Experts	test
1	20	25	Marzetta Thomas L.			
2	18	18	Kim Youn Sun			
3	18	19	Lee Ju Ho			
4	17	20	Ashikhmin Alexei			
5	17	17	Lee Hyo Ji			
6	14	14	CHO Young			
7	13	13	L Louay			
8	13					
9			SHER Amin			
			K Sam			
16	8	9	Ahn Jae Young			
17	8	8	LEE Kilbom			
18	8	9	Nam Jun Young			
19	8	8	OKUMURA Yukihiro			
20	8	8	SUYAMA Satoshi			
21	8	10	Yang Hong			

Thomson Data Analyzer Analyst's Guide (online)

News

①选中需要导出publication no.的项(比如 Ashikhmin Alexei), ②然后, Scripts-> Export-Copy Publication Numbers to Clipboard

Scripts Window Help

Run Script...  
Modify Script Menu...

Report - Company Comparison Ctrl+1  
Report - Technology Report Ctrl+2  
Report - Company Report Ctrl+3  
Report - TDA Pivot Charts  
Report - Patent Vital Signs

Clean - DWPI Cleanup (DWPI only) Ctrl+9  
Report - Basic Report (DWPI only)  
Report - IPC based Analysis (DWPI only)  
Report - Top Assignees (DWPI only)  
Report - Trend Analysis (DWPI only)

Clean - Combine Author Networks

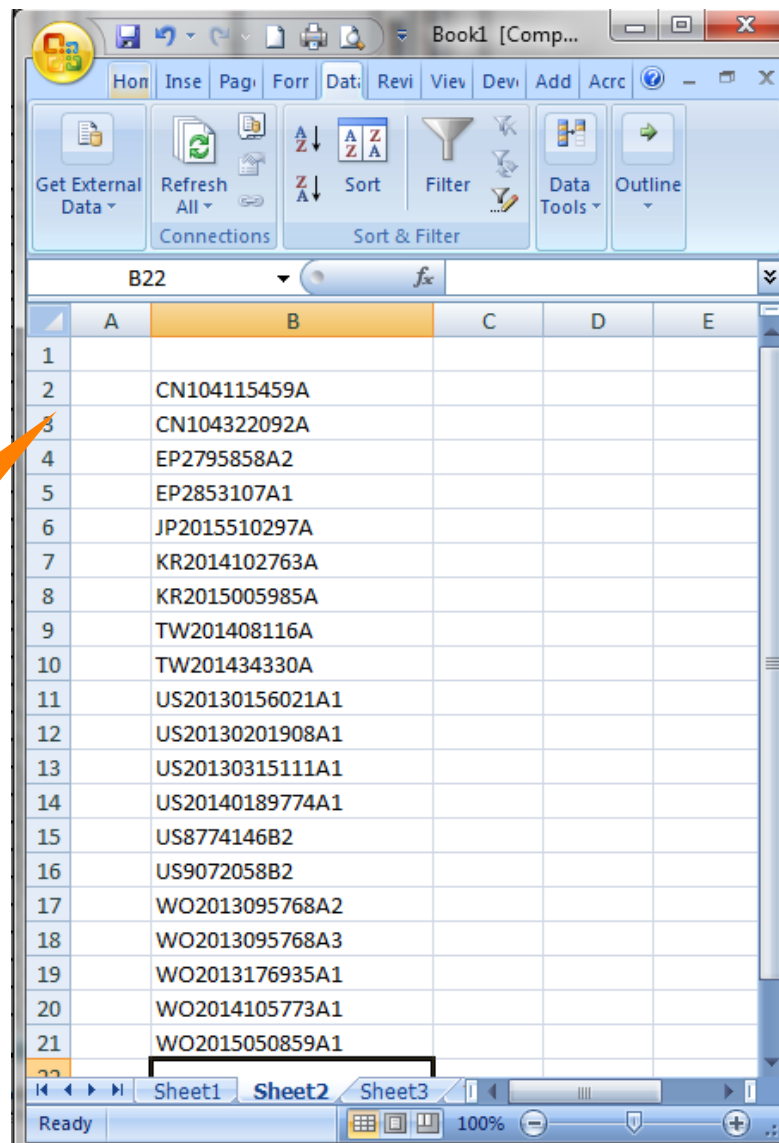
Report - Aduna Cluster Map  
Report - Bubble Chart  
Report - Company Activity Gantt  
Report - Make Pivot Chart in Excel  
Report - Plot List in Excel  
Report - Plot Matrix in Excel  
Report - Super Profile  
Report - Terms by Year  
Report - World Map  
Report - Plot Classification  
Report - Browse Classification Tree

Export - Copy Publication Numbers to Clipboard

Utility - AND search  
Utility - Concatenate Fields  
Utility - Create All Lists  
Utility - Combine Groups  
Utility - Make Thesaurus

## Export-Copy Publication Numbers to clipboard

③ 复制到excel中



The screenshot shows an Excel spreadsheet with the following data in column B:

	A	B	C	D	E
1					
2		CN104115459A			
3		CN104322092A			
4		EP2795858A2			
5		EP2853107A1			
6		JP2015510297A			
7		KR2014102763A			
8		KR2015005985A			
9		TW201408116A			
10		TW201434330A			
11		US20130156021A1			
12		US20130201908A1			
13		US20130315111A1			
14		US20140189774A1			
15		US8774146B2			
16		US9072058B2			
17		WO2013095768A2			
18		WO2013095768A3			
19		WO2013176935A1			
20		WO2014105773A1			
21		WO2015050859A1			

# Utility-combine Groups

	# Records	# Instances	1st author's	test1	test 2	test 3	all
1	3	3	Beijing Forestry Univ, MOE Key Lab Wooden Mat Sci & Applicat, Beijing 100083, Peoples R China Liu, Yi	☑			☑
2	2	2	Beijing Forestry Univ, Beijing Key Lab Lignocellulos Chem, Beijing 100083, Peoples R China Dai, Lin	☑			☑
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		☑		☑
13	1	1	Beijing Agr Univ, Food Sci & Engrn Coll, Beijing 100094, Peoples R China Zhang, Yanyan		☑		☑
14	1	1	Beijing Forestry Univ, Beijing 100083, Peoples R China Lu, Nan		☑		☑
15	1	1	Beijing Forestry Univ, Beijing Key Lab Forest Food Proc & Safety, Coll Biol Sci & Technol, Dept Food Sci & Engrn, Beijing 100083, Peoples R China				
16	1	1	Beijing Forestry Univ, Beijing Key Lab Forest Pest Control, Beijing 100083, Peoples R China Zhang, Lu				
17	1	1	Beijing Forestry Univ, Beijing Key Lab Lignocellulos Chem, Beijing 100083, Peoples R China Chen, Chang-Zhou				
18	1	1	Beijing Forestry Univ, Beijing Key Lab Lignocellulos Chem, Beijing 100083, Peoples R China Chen, Jinhui				
19	1	1	Beijing Forestry Univ, Beijing Key Lab Lignocellulos Chem, Beijing 100083, Peoples R China Chen, Jinhui			☑	☑
20	1	1	Beijing Forestry Univ, Beijing Key Lab Lignocellulos Chem, Beijing 100083, Peoples R China Chen, Jinhui			☑	☑
21	1	1	Beijing Forestry Univ, Beijing Key Lab Lignocellulos Chem, Beijing 100083, Peoples R China Chen, Jinhui			☑	☑
22	1	1	Beijing Forestry Univ, Beijing Key Lab Lignocellulos Chem, Beijing 100083, Peoples R China Chen, Jinhui			☑	☑
23	1	1	Beijing Forestry Univ, Beijing Key Lab Lignocellulos Chem, Beijing 100083, Peoples R China Chen, Jinhui			☑	☑

将某个字段中的所有组全部添加到新组中，利用Utility-Combine Groups。比如图中第三组是对test1和testB的融合

# 矩阵

生成矩阵

行

- Abstract
- Abstract (Best Available)
- Abstract (Derwent)
- Abstract (Derwent) (NLP) (Phrases)
- Abstract (Derwent) (NLP) (Words)
- Abstract (NLP) (Phrases)
- Abstract (NLP) (Words)
- Abstract ADVANTAGE
- Abstract ADVANTAGE (NLP) (Phrases)
- Abstract ADVANTAGE (NLP) (Words)
- Abstract DESCRIPTION OF DRAWINGS
- Abstract DESCRIPTION OF DRAWINGS (NLP) (Phrases)
- Abstract DESCRIPTION OF DRAWINGS (NLP) (Words)
- Abstract DETAILED DESCRIPTION

列

- Abstract
- Abstract (Best Available)
- Abstract (Derwent)
- Abstract (Derwent) (NLP) (Phrases)
- Abstract (Derwent) (NLP) (Words)
- Abstract (NLP) (Phrases)
- Abstract (NLP) (Words)
- Abstract ADVANTAGE
- Abstract ADVANTAGE (NLP) (Phrases)
- Abstract ADVANTAGE (NLP) (Words)
- Abstract DESCRIPTION OF DRAWINGS
- Abstract DESCRIPTION OF DRAWINGS (NLP) (Phrases)
- Abstract DESCRIPTION OF DRAWINGS (NLP) (Words)
- Abstract DETAILED DESCRIPTION

矩阵：共现（基于记录数量）  
190 行：Abstract --全部项  
0 列：无字段选定

矩阵类型

- ☒ 共现
- ☐ 自相关
- ☐ 交互相关

基于

- ☒ 记录数目
- ☐ 出现频次

范围

- ☒ 记录
- ☐ 源项

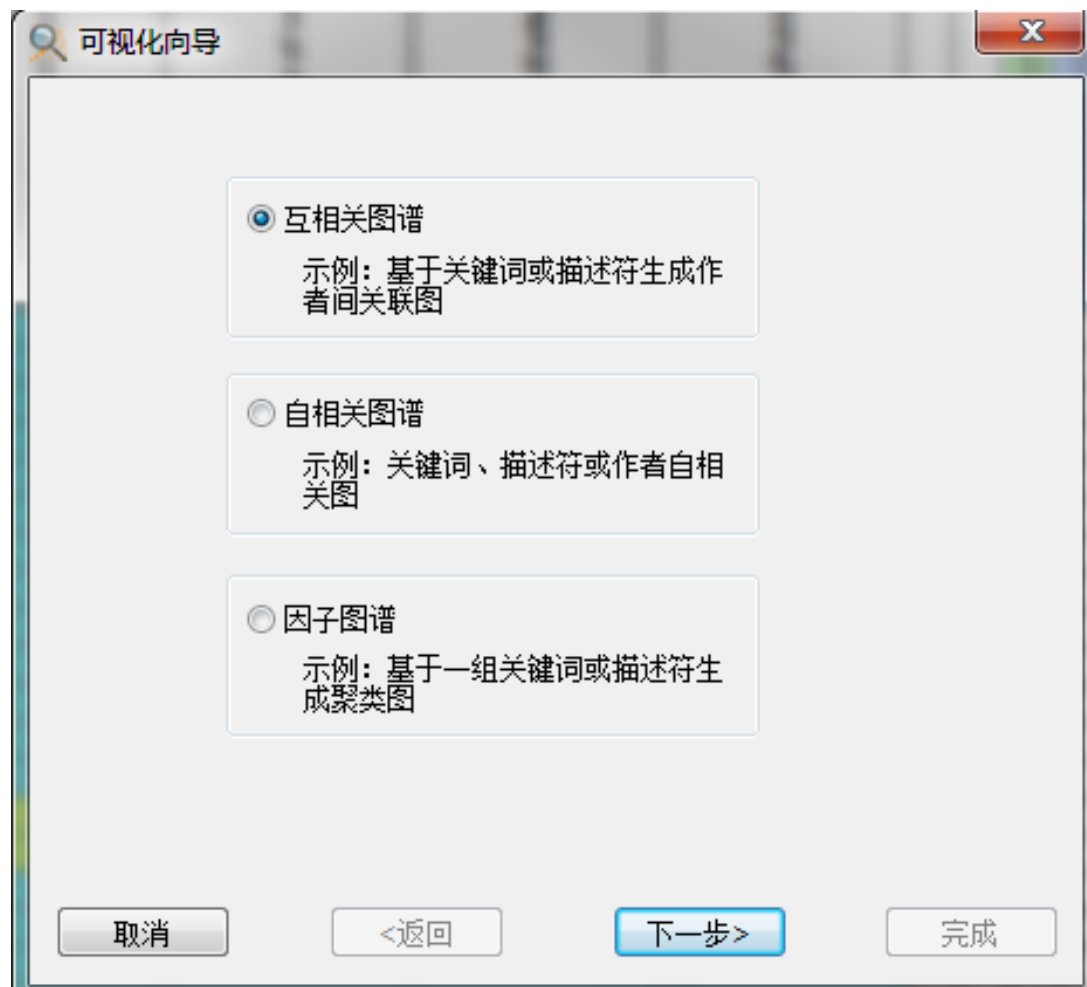
相关函数

- ☒ 皮尔逊系数
- ☐ 余弦
- ☐ 最大比例

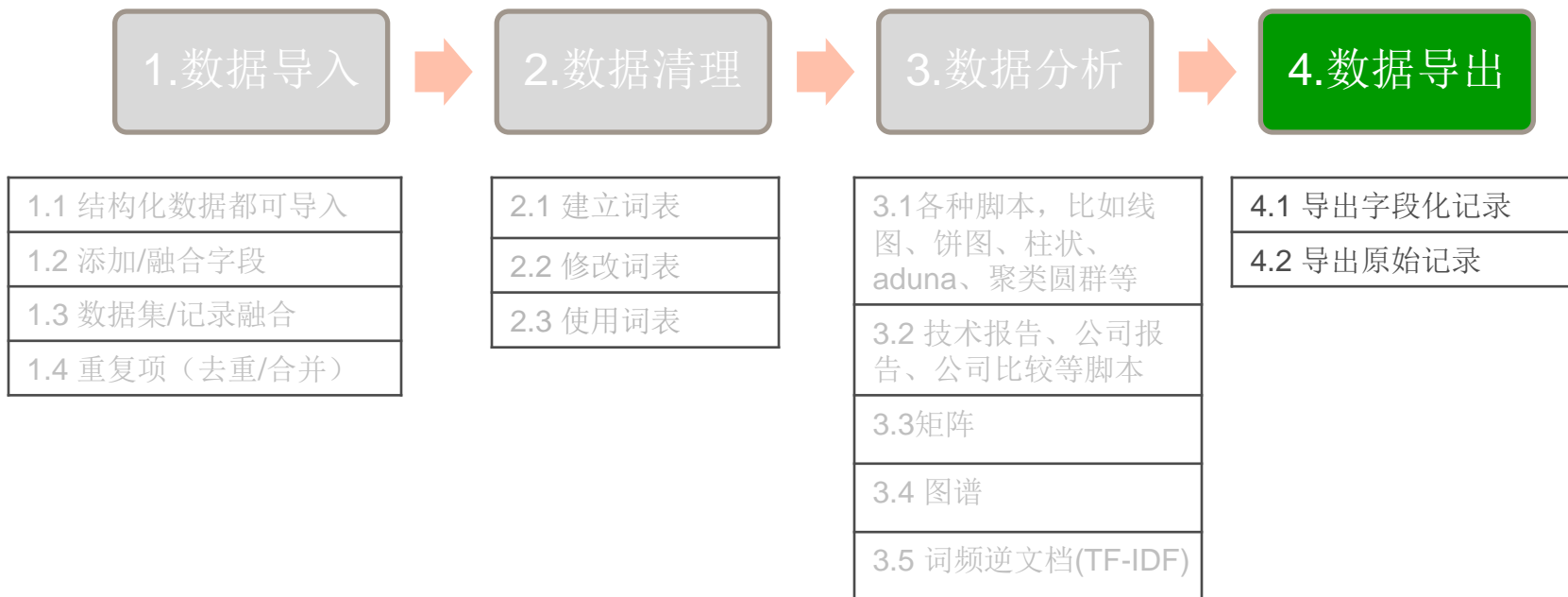
帮助 确定 取消



## 图谱:



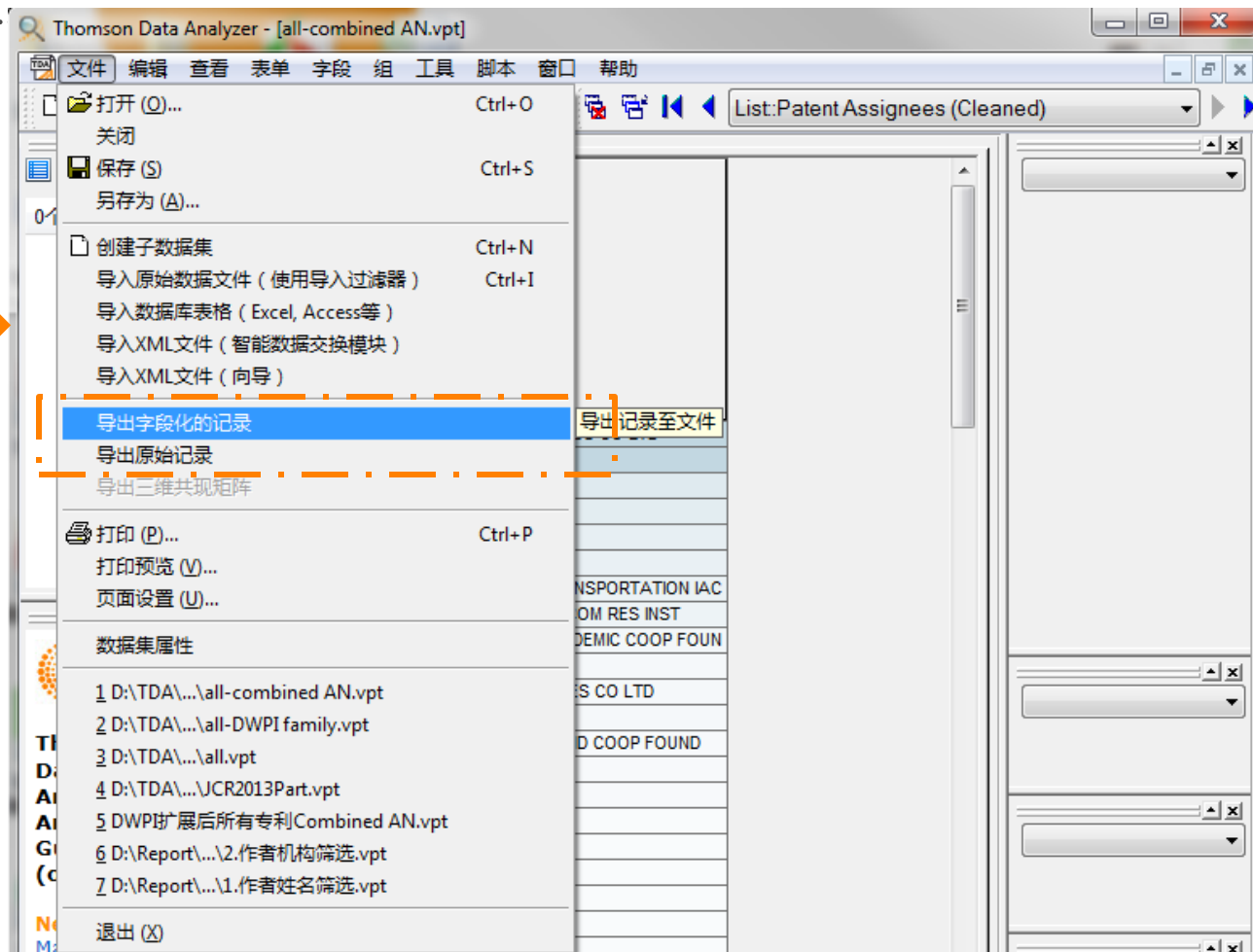
# TDA简要流程概览



## 4.数据导出

4.1 导出字段化记录

4.2 导出原始记录



THOMSON REUTERS



# 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  |





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