

中国科学院国家科学图书馆
公派访问学者交流学习总结

总结人：胡正银

2012.2.28

一、访问目的.....	3
二、学术业务研究收获.....	3
1. 研究团队情况介绍.....	3
2. Term Clumping 研究.....	4
3. 基于 Semantic TRIZ 的专利文献 Problems & Solutions 建模研究.....	7
4. 技术挖掘方法及软件、工具学习.....	7
5. 参加会议及学术报告情况.....	8
6. 项目组工作方式.....	10
三、对相应业务工作的改革创新设想.....	11
1. 数据、知识积累.....	11
2. 多学科交流机制.....	12
3. 方法、工具的学习与交流.....	12
四、学习成果.....	12

我得到中国科学院国家科学图书馆群星计划资助，于 2011 年 8 月—2012 年 2 月作为访问学者赴美国佐治亚理工学院公共政策学院科技政策与评估中心（TPAC）进行了为期半年的访问研究和学习，并于 2012 年 2 月 18 日按期回国。

一、访问目的

我这次访问的主要目的是学习佐治亚理工公共政策学院 TPAC 在技术挖掘方面的研究和应用情况，考察和学习佐治亚理工公共政策学院在科研及管理方面的优点和先进经验，兼感受美国科研文化。

二、学术业务研究收获

1. 研究团队情况介绍

在访问期间，我参加了 Philip Shapira 教授和 Jan Youtie 教授领导的“纳米技术与创新体制评估”（Nanotechnology Research and Innovation Systems Assessment）项目中“染料敏化太阳能电池”（Dye-sensitized solar cell, DSSC）工作。团队成员包括四名资深研究人员（Philip Shapira 教授、Jan Youtie 博士（项目主管）、Alan L Porter 教授、Juan Rogers 教授）、一名博士后、四名博士、一名硕士，其专业方向涉及：科技政策与评估、经济学、技术挖掘、计算机软件、文献计量学等。访学期间，该团队一共有 3 位来自中国长期访问学者/访问学生，接待过来自哥伦比亚、巴西、意大利 3 位短期访问学者。



图 1 项目组部分成员合影

该项目的研究重点包括以下三个方面：

- 纳米技术研究和发展的概况——探索纳米技术的研究方向，主要科研机构，及其研究网络和关系。
- 纳米技术的核心及前沿地区——识别纳米发展的重要及新兴地区。
- 纳米产业化——纳米产业评估，包括纳米公司，风险投资，纳米应用和纳米产品。

佐治亚理工学院的研究团队隶属于纳米技术和社会研究中心(CNS-ASU)。CNS-ASU 位于亚历桑那大学，成员包括佐治亚理工和亚利桑那大学在内的六所美国著名大学，并获得美国自然科学基金委员会(NSF)的资助，以从事纳米技术发展的研究。佐治亚理工学院致力于在新世纪中纳米技术反思和社会影响的探讨。

在 2012 年 1 月，课题组又获得英国经济与社会研究委员会(Economic & Social Research Council)资助，以绿色技术为主要研究对象，开展创新型企业可持续发展研究(Sustaining Growth for Innovative New Enterprises)。

2. Term Clumping 研究

“Term Clumping”是 Alan Porter 教授提出来的一个概念，主要内容包括：目

标文档集进行主题术语抽取、清洗、合并、语义分析等。通过 Term Clumping 形成的主题术语集可进行面向未来的技术分析。Alan Porter 教授定义的”Term Clumping”框架图如下：

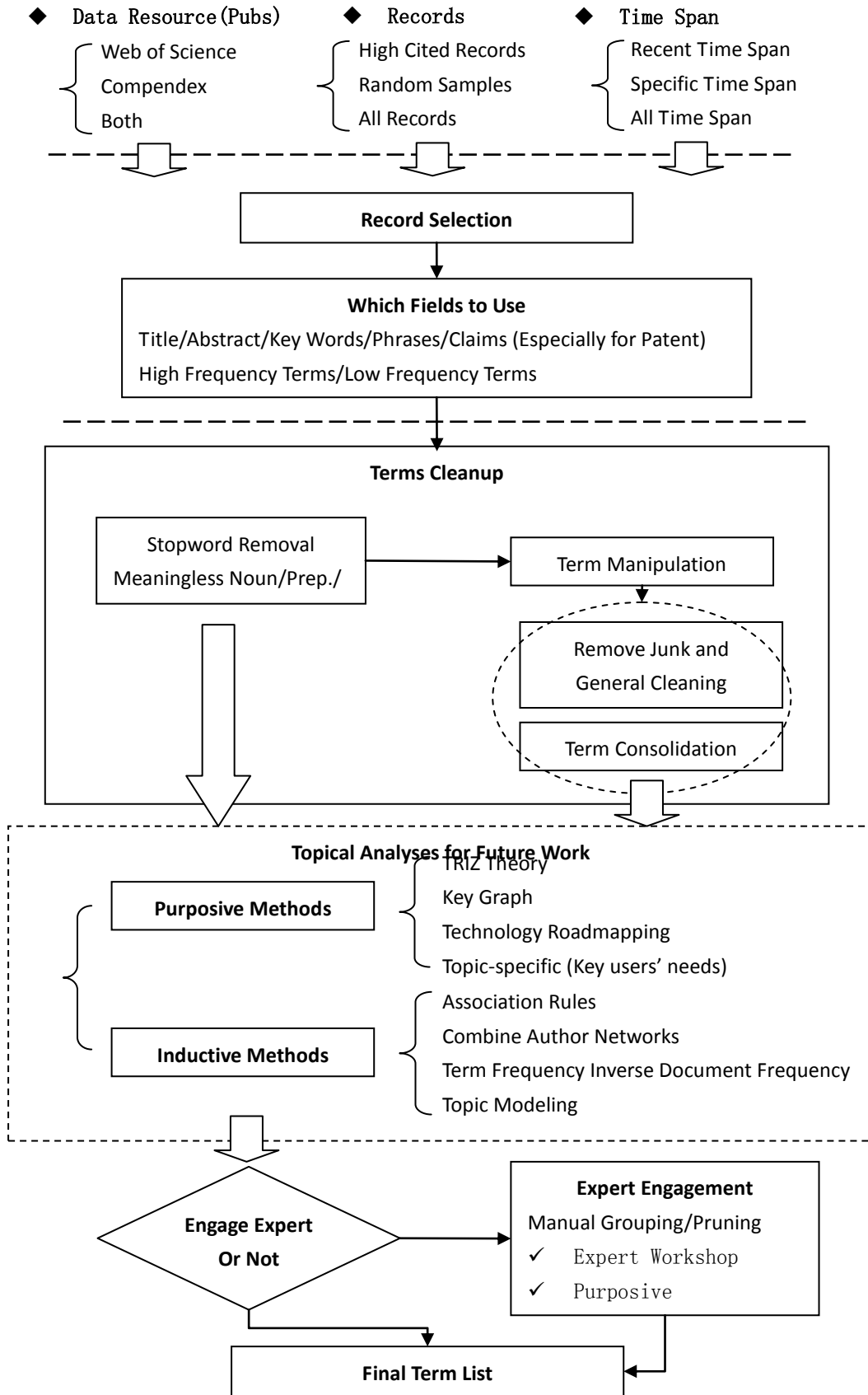


图 2 Term Clumping 框架

在佐治亚理工访问期间，在 Alan Porter 教授指导下参与了 Term Clumping 技术调研与实践。并以 DSSC 为目标数据集，开展研究工作。调研了 Term Clumping 中涉及到的较新的分析方法，如：KeyGraph、Topic Model、Semantic TRIZ 等技术主题监测与建模方法。我负责从技术上实现了 KeyGraph 技术主题监测方法，编写 KeyGraph 与技术挖掘软件 VantagePoint (TDA 原始版本) 数据交换的接口。用 KeyGraph 对 DSSC 数据进行主题监测分析，与 Topic Model 及传统 PCA(Principal components analysis)分析结果展开比较。图 3 是利用 KeyGraph 处理 DSSC 数据结果截图。

```

KEYWORDS:      data storage equipment, chemical properties, cadmium, cadmium compounds,
DOCUMENTS:     1330,
               597,
               1125,
KEYGRAPH_NODES: 234: data storage equip: data storage equipment, 2173: chemi c properti: chemical properties, 561: cadmium: cadmium, 233: cadmium
KEYGRAPH_EDGES:
E173: chemi c properti-234: data storage equip, 561: cadmium-234: data storage equip, 233: cadmium compound-234: data storage equip,
KEYWORDS:      radical cation, resolved absorption spectroscopy, retinoic acid, time resolved spectroscopy, singlet triplet annihilation
DOCUMENTS:     4331,
               5266,
KEYGRAPH_NODES: 2300: radic cation: radical cation, 2298: resolv absorp spectroscopi: resolved absorption spectroscopy, 5906: retino acid: re
KEYGRAPH_EDGES:
E300: radic cation-4640: singlet triplet annihil, 2300: radic cation-1833: squarylium cyanin, 2298: resolv absorp spectroscopi-5906: retino a
KEYWORDS:      thin film electrode, surface charge, solar absorbers, semiconductor interfaces, space charge capacitance, time resolved ph
DOCUMENTS:     1797,
KEYGRAPH_NODES: 4260: thin film electro d: thin film electrode, 2512: surfac charg: surface charge, 1145: solar absorb: solar absorbers, 4353: s
KEYGRAPH_EDGES:
E260: thin film electro d-2511: time resolv photocurr, 2512: surfac charg-2511: time resolv photocurr, 1145: solar absorb-2511: time resolv ph
KEYWORDS:      tripod sensitizers, spectral properties, reaction intermediates, semiconductor nanoparticles, rigid rod linkers,
DOCUMENTS:     1500,
KEYGRAPH_NODES: 4102: tripod sensit: tripod sensitizers, 2955: spectral properti: spectral properties, 6658: reaction intermedi: reaction i
KEYGRAPH_EDGES:
E955: spectral properti-4102: tripod sensit, 6658: reaction intermedi-5550: semiconductor particl, 4103: rigid rod linker-5550: semiconductor
KEYWORDS:      sensitized nanocrystalline tio2, rhenium, relaxation dynamics, relaxation processes,
DOCUMENTS:     3274,

```

图 3 KeyGraph 处理 DSSC 数据结果截图

3. 基于 Semantic TRIZ 的专利文献 Problems & Solutions 建模研究

访学期间，与导师及同事广泛探讨了基于 Semantic TRIZ 的专利文献 Problems & Solutions 建模研究。Alan Porter 教授除了提供一篇极具针对性的参考文献“Identifying technology trends for R&D planning using TRIZ and text mining”外，还联系西班牙一位 TRIZ 专家与我们一起讨论。在美期间，初步完成“Construction Problem & Solution Models of Patent Documents”方案。拟回国后针对具体领域开展相关实证研究。图 4 是我在学术交流上所做关于专利文献 P&S 建模 PPT 讨论大纲。



Construction Problem & Solution Models of Patent Documents

- Why?
- How to build Model Schema?
- How to get Model Instances?
- Application Scenario of models
- Evaluation of Model or Empirical Analysis

图4 专利文献 P&S 建模讨论 PPT 纲要截图

4. 技术挖掘方法及软件、工具学习

访问期间，参加了课题组组织关于技术挖掘方法、软件工具学习，具体如下：

TPAC 的技术挖掘方法主要有 10 步骤，如图 5 所示。



- 
- #### 2. Ten-Step Tech Mining
1. Spell out the Intelligence questions and how to answer them
 2. Get suitable data
 3. Search (iterate) & retrieve ~abstract records
 4. Import into text mining software (VantagePoint)
 5. Clean the data
 6. Analyze
 7. Visualize (Map)
 8. Integrate with Internet analyses & expert opinion
 9. Summarize; Interpret; Communicate (multi-dimensionally)!
 10. Standardize and semi-automate where possible
- 

图5 技术挖掘 10 步骤

TPAC 开展文献计量分析的工具主要有：

- VantagePoint: 由 Alan L Porter 教授于 1996 年左右开发并商业化, TDA 是由汤姆森公司商业化的 VP 版本。Alan 教授的 Seach Technology 公司一直在做后续开发,并针对不同的数据库编写可用 VP 进行分析的 filter。通过培训,更深入了解 VantagePoint 的高级应用技术,如: Macros、叙词表应用,数据可视化聚类分析等。
- Pajek: 免费的可视化软件,主要用于制作科学地图。

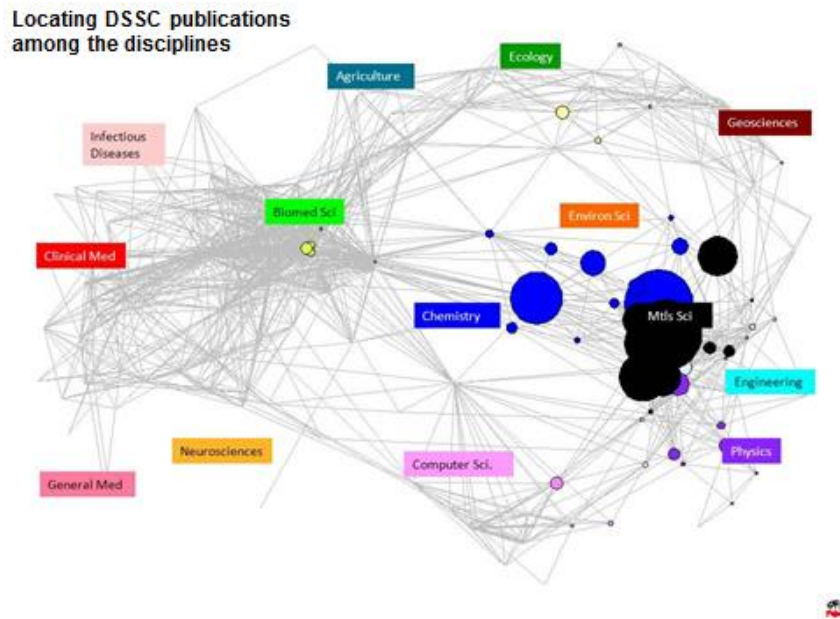


图 6 Pajek 制作的科学地图

- Gephi: 开源免费软件,可用来可视化各种图及关系。Gephi 还提供众多的 API 供开发人员使用,在课题组应用十分广泛。项目组的博士后 Luciano 开发出一个脚本 (script),可将 VP 数据导入 Gephi,进行可视化。

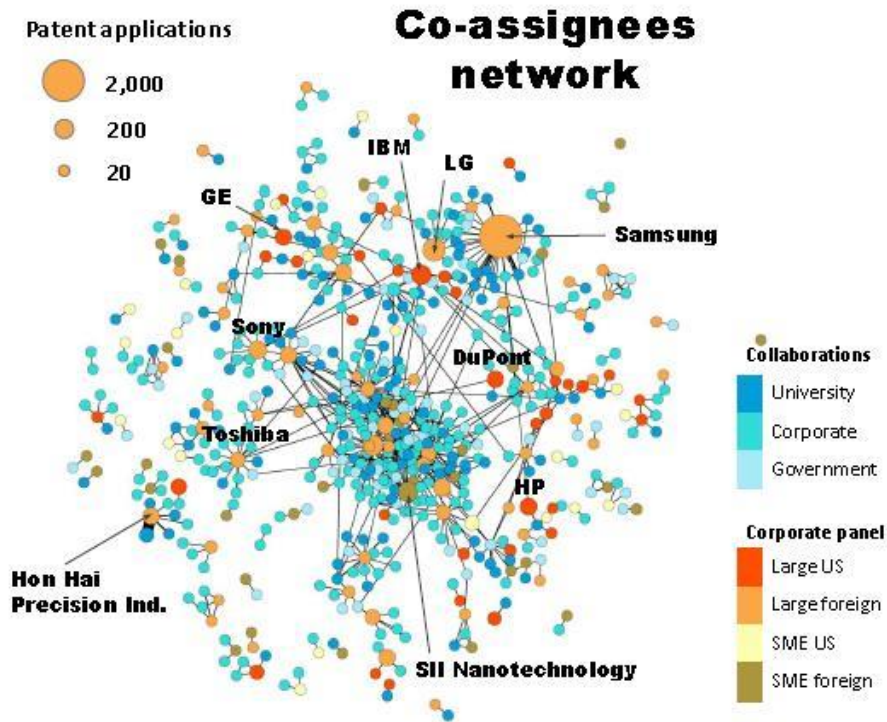


图7 Gephi制作的专利权人合作图

5. 参加会议及学术报告情况

访问期间,除了参加课题组小组例会外,还参加了一些其他会议与学术报告,开阔了视野,具体如下:

- 1ST Global TechMining Conference:参加了 Alan Porter 教授组织发起的 1ST Global TechMining Conference, 该会议于 2011 年 9 月在佐治亚理工学院召开。在会上做了“Analyzing Technology Evolution of Graphene Sensor Based on Patent Documents”的口头报告。
- 参加一次佐治亚理工公共政策学院与美联储亚特兰大分行联合举办的经济发展创新论坛(Innovations in Economic Development Forum)。听取美国国家科学院、美联储相关专家介绍科技创新对美国经济的推动作用及目前美国面临的全球性挑战。
- 参加多次佐治亚理工学院华人学者论坛,听取了包括:佐治亚理工学院杰出教授,中国科学院外籍院士王中林教授、佐治亚理工学院和埃默里大学生物医学工程系教授,美国纳米医学研究所所长包刚教授在内众多

学者关于科研、人生、理想的讲座。

6. 项目组工作方式

项目组自建有一个工作平台，上传所有可共享的资料和项目进展相关材料。团队成员均可上传和下载相关信息。项目组专用服务器上有关于技术挖掘软件、分析脚本、测试数据等资料供下载。此外，项目组成员发现好的文章或者资源，一般会通过邮件列表发送给项目组所有成员共享。

项目组分工一般由导师或项目负责人根据各人专业分配，各组员独立或协作完成各自工作。每周五上午 10-12 点，整个团队会举行一次例会，一般会议由 Jan 主持，若 Jan 出差，则由其他研究人员主持。主持人在每次会议开始前会先列出本次会议需要重点讨论或汇报的内容，各成员补充。每名成员自愿作自己阶段性工作的汇报。成员也可以在会上提出自己在科研中遇到的问题向大家请教，或者就自己的论文草稿征求项目组其他成员的意见。主持人会根据议题，不定期邀请其他领域专家参与交流讨论，比如：邀请佐治亚理工学院纳米科学专家、计算机科学专家进行交流。会议气氛非常轻松愉快，成员随时可以根据自己的时间表离开（主持人不会）。遇到不感兴趣的话题，可以打开笔记本或者 IPAD 做自己的工作。

此外，项目组与北京理工大学管理与经济学院朱东华教授课题组、英国曼切斯特大学商学院建立有 Co-Lab 交流机制。会不定期利用网络举行三方视频会议，进行交流。

三、对相应业务工作的改革创新设想

1. 数据、知识积累

TPAC 课题组有良好的数据及知识积累制度与平台。包括：全球纳米数据、检索策略、数据清洗规则、VantagePoint 针对特定分析的 Macros、词表等都得到很好的积累与保存。其他成员要共享这些资源非常方便。

我馆每年会开展多项专利分析或学科分析，也应当考虑将各个领域的数据、知识进行统一存储与管理，以便于其他用户共享，提升研究效率。特别对于数据

检索来说，如果在先研究的检索策略和检索结果能够保存起来，并可被其他研究人员获得，那其他研究人员在开展相同领域的研究时，可以在先前研究的基础上作进一步完善检索策略，提高工作效率。

2. 多学科交流机制

TPAC 课题组成员学科背景及其多样化，项目组长 Phil 是商学院教授，主要研究企业创新政策，Jan 则是经济学院教授，在利用经济学理论方法进行科技评估分析方面有独到见解，Alan 这是技术挖掘方面的领军人物。项目组博士生中既有文献计量学背景，也有计算机背景，还有科技政策背景等。每次在一起讨论，都能感觉对方谈话内容中带有鲜明的学术背景烙印。

我馆同事学科背景也极为多样化，是否也可以组建跨总分馆、跨部门、跨研究方向的多学科交流机制？比如：不定期组织我馆不同部门同事就某一主题，从自己的专业或者管理角度，发表自己的看法。或者就自己工作中遇到的疑难提出来，请不同专业背景的同事提些参考建议？跨学科的脑力风暴，往往更能激发创新的火花。

3. 方法、工具的学习与交流

Alan Porter 教授的技术挖掘 10 步骤，对技术预见分析做了很好的总结了，不同学科背景的同事都可从中受益。

VantagePoint 与 TDA 同源同质，在 Search Technology 公司网站上有大量关于导入 Filter、词表、高级分析 Macros 下载。如果有特定分析需求，还可以给公司技术支持团队发邮件，请求技术支持。

Pajek、Gephi 在数据可视化方面大有可为，尤其是 Gephi 功能及其强大，号称是 Photoshop for Graph，可以让有兴趣的同事深入学习后，在全馆范围内进行交流培训。

四、学习成果

- Yi Zhang, Alan Porter, Zhengyin Hu (2012). An Inductive Method for “Term

Clumping²: A Case Study on Dye-Sensitized Solar Cells. The International Conference on Innovative Methods for Innovation Management and Policy (IM2012), Beijing.

- Study on construction problem & solution models of patent documents.(unfinished).