

从“科学之科学”^①到“政策之科学”

□ 治部真理 山崎雅和 (原著) / 日本科学技术振兴机构 日本东京 102-8666
 李颖 (编译) / 中国科学技术信息研究所 北京 100038

摘要: 由于纵观科学结构的所谓“科学之科学”(Science of Science)正转向“科学政策之科学”(Science of Policy)的新方向,为此可以认为,纵观科学结构的科学地图(science map)将移向科学结构中包含的纵览政策的政策地图(policy map)方向。在此背景下,基于科学根据(evidence)的政策制定可以确保其透明性。文章通过应用日本科学技术振兴机构(简称JST)构建的专利和论文整合检索系统(PATLISYS-J),介绍世界科学结构的可视化,以及将任意研究者详细分析等进行可视化的方法。

关键词: 政策科学, 政策地图, 科学地图, Cytoscape, 专利和期刊论文, 链接

DOI: 10.3772/j.issn.1673-2286.2011.03.005

1 简介

1.1 背景

日本在2001年4月15日举办的第七届综合科学技术会议基本政策专门调查会上,公布了“科学技术基本政策制定的基本方针(草案)”。据此,为了强化从基础研究到创新的各个环节,国家要求建立“活用学术论文和专利数据库、基于府省共通研究开发管理系统(e-Rad)的政府研究开发数据库,按照不同机构或不同研究领域,明确大学及研究开发机构的资金投入和产出间的关系,反映资金分配上的计划”这样的体系,在促进“专利和论文信息整合系统应用的同时,完善相关专利及各种文献间链接和分析系统,促进知识产权关联信息的基础整备和网络化”^[1]。

除此之外,作为“新政策的展开”工作,要求“实施基于科学证据(evidence)的政策制定”,“促进‘政策之科学’,确立完备的评价指标及政策效果的分析方法”^[2]。

溯源历史性的科学定量评价方法,该分析手法的基础构建始于普赖斯(Price D)^[3]。基于此,1972年美国国家科学基金会NSF发布了“Science and Engineering Indicators”(科学与工程指标)。各国也随后陆续发表了“Science and Engineering”的指标。日本则发表了“科学技术指标: Science and Technology Indicators”。在日本,有关“科学技术”这一术语利用的原委,在“科学技术厅政策史——其成立和发展”中给予了详情的总结,可以参照参考文献^[4]。另外,在下文出现的英

语标记中,比如Map of Science、Science mapping、Atlas of Science,以及“科学政策之科学: Science of Science Policy (SoSP)”^[5]中的‘Science’,对日本而言,可认为“科学技术”包含其中。

其后,各国政府开始采用定量手段分析科学的发展进步,利用这些指标开创新领域、对哪些领域应当配以多大的预算等等,都急需进行客观的判断决定。

在社会主义计划时代旧苏联体制中,从国家“包含人才的研究开发投入资源和由此取得的成果,即投入(资源)和产出(成果)之间达到最适宜平衡”的需要出发,就着眼于普赖斯的方法论,进行了相关分析^[6]。

确立科学技术指标,实现制度化,在“科学计量学”历史发展过程中,还需提及“文献计量

^①Science of science: 《科学》杂志 (Science Magazine) 的卷首语。

学”的历史。

文献计量学的历史在很大程度上依存于尤金·加菲尔德 (Eugene Garfield)。他在1950年创立了ISI (Institute for Scientific Information, 美国科技情报所), 推出了“current contents”和“SCI”。同时, 开发了吸引馆员兴趣的文献检索工具, 以引文分析为切入点, 发表了“最多被引用科学家 (citation superstar, most cited authors)”、“最多被引用领域 (Research fronts, most cited subject areas)”、“地图 (map, science graph)”^[7]。这些引用分析可以说是基于庞大司法判例的引用关系进行整理方法的应用。1981年完成了“科学地图”的实验版本, 商品化为“Atlas of Science” (科学地图)。加菲尔德销售计算机软件“SCI-MAP”是在1989年, 在当今的“Science Watch” (科学观察) 中, 作为“Hot Paper”被予以介绍^[8]。

1992年, 汤姆森公司收购了ISI, 成为汤姆森集团 (The Thomson Corporation) 旗下的子公司。加菲尔德开发的各种软件包, 如“ISI Web of Knowledge”、“Science Citation Index”、“Essential Science Indicators”中的“Research Front” (研究前沿)、“Hot Papers” (热点论文) 等等, 一直在延续。另外, “Atlas of Science”在“Science Watch”中成为了“Research Front Map” (研究前沿地图)。

到上世纪90年代, 国际科学计量学和情报计量学学会创立之后, 在文献计量学和计量科学的融合过

程中, “科学具有结构”这一科学之科学的分析开始盛行。

从“科学之科学: Science of Science”到“科学政策之科学: Science of Science Policy”转向的契机是前布什政权时代的科学技术政策局 (OSTP: Office of Science and Technology Policy) 局长兼总统科学顾问John Murberger三世于2005年4月, 在美国科学振兴协会 (AAAS) 科学技术政策论坛上的主旨演讲。该主旨演讲中, Murberger倡导“联邦政府向研究开发投资, 在科学政策制定之际, 为支持科学政策担当者而构建数据集、工具、创出方法论的实践社区”。

Murberger也在2006年渥太华召开的OECD Blue Sky II Forum中, 进行了主旨演讲。此会议是为了明确政策需求、指标测定问题、创新领域中跨学科新主题而举办的。第一届Blue Sky Forum于1996年在巴黎召开, 仅局限于强调科学技术指标 (Science and Technology Indicators) 的重要性^[8], 第二届强调了“与政策需求挂钩”科学和技术及创新指标 (Science, Technology and Innovation Indicators) 的重要性^[9]。

在日本, 内阁府、日本经济团体联合会、日本学术会议、GIES 2007组织委员会共同主办了GIES 2007 (global innovation ecosystem 2007)^[2]和GIES2008^[3], 设置了“评价测定框架构建”的分会。那时开始, 由于日本经济成长的困难, 连续严峻的政府财政问题, 针对政府科学技术上的投入, 应当具有怎样的效益? 政府被追究要承担说明的

责任和其投入的效益。

1.2 从科学地图学转向科学技术政策地图

世界上最初开发科学地图者为ISI公司的Henry Small^[10], Small于1973年采用引文分析、尤其是共引分析 (Co-citation analysis) 来把握论文的关联性, 将其地图化。现在, 汤姆森公司的Research Front^[4]及爱思维尔公司的SciVal^[5]等, 从商业产品到计量科学及文献计量学研究者开发的系统多数存在。从1973年以100篇论文为对象进行分析和地图化, 现在已经发展为以TB数量级的论文及专利为对象进行分析和地图化。

图1为1999年发表的有关1996年的科学地图。这是基于1996年SCI数据, 利用共引分析作出的图形^[6]。

利用引文的另一种分析方法是文献耦合分析 (bibliographic coupling analysis)。这是1963年由Kessler提倡的^[11]。共引分析 (Co-Citation Analysis) 是以同时被第3者论文引用的论文为对象的聚类分析 (cluster analysis), 与此对应, 文献耦合分析以共有引用对象论文为对象。前者伴随引用时间的变化同时结构会变化, 后者可看到不随时间变化的不变结构。目前共引分析频繁地被利用, 与此对比, 文献耦合分析几乎不再被利用。

如图2所示, 现在存在着多种多样形态的科学地图, 这些可以在<http://www.scimaps.org>上浏览到。另

^[2]有关GIES2007, 参照<http://crds.jst.go.jp/GIES/archive/GIES2007/>

^[3]有关GIES2008, 参照<http://crds.jst.go.jp/GIES/archive/GIES2008/>

^[4]参照Science Watch (<http://sciencewatch.com/>)

^[5]参照<http://www.scival.com/>

^[6]<http://www.scimaps.org> 1.5 1996 Map of Science: A Network Representation of the 43 Fourth Level Clusters Based on Data from the 1996 Science Citation Index, by Henry Small Philadelphia, PA 1999 Courtesy of Henry

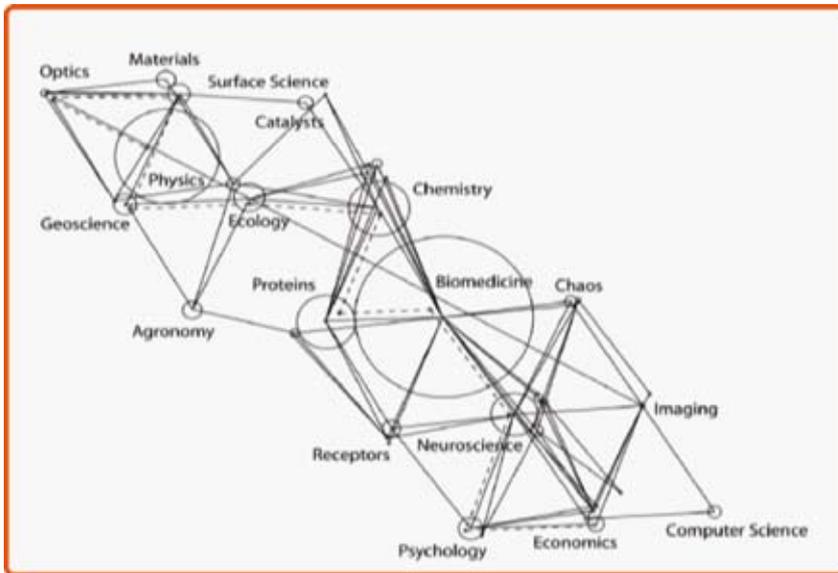


图1 Henry Small的科学地图^①



图2 多种多样的科学地图

外，印第安纳大学Katy Borner的著作“Atlas of Science Visualizing What We Know”计划进行汇总^[12]。

从纵观科学结构的科学地图变化到科学政策的地图时间点大约

为2009年左右。图3是针对2007年接受NIH（美国国立卫生研究院）资助的研究者在申请基金时所用文摘进行了词素解析^②之后，将词汇共现关系聚类而形成的地图。

这和在“信息管理杂志”上介绍的以JST事业可视化为例的手法相同^[13,14]。如前所述，通过关联基于府省共通研究开发管理系统（e-Rad）的政府研究开发数据库和论文数据库以及专利数据库，如同NIH的政策地图一样，JST等的各个基金机构、或者每一竞争性资金制度，通过将科学技术结构的地图化，可构建支持资金分配的系统，可以认为，这是朝向实现基于科学证据（evidence）的政策制定迈出的一步。

在这种背景下，本文利用独立行政法人科学技术振兴机构（以下简称JST）的论文和专利整合检索系统（以下简称PATLISYS-J）^[15]，介绍作为政策之科学的“政策地图”。

2 分析方法

2.1 数据库

PATLISYS-J中收录的科技文献及专利文献如下：

论文：Elsevier公司的Scopus Custom Data，从1996年到2007年，约15000万件

专利：EPO PATSTAT（Worldwide Patent Statistical Database）800万件

Elsevier公司的Scopus Custom Data收录了世界5000多个出版社的18000多种杂志，3800多万件的书目、文摘记录。

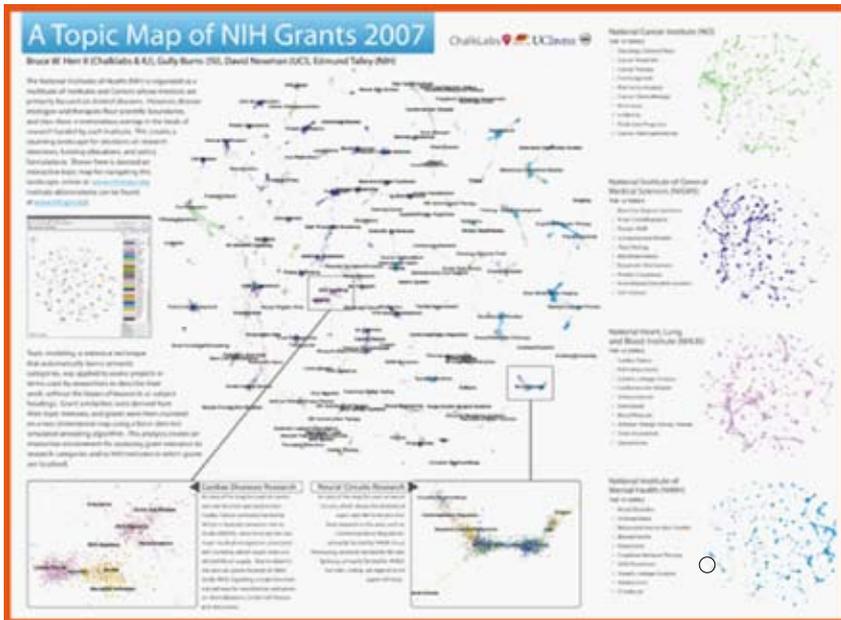
2.2 人名称的合并

Small, Thomson Scientific ISI

^①参照<http://www.sciaps.org>。

^②词素解析：把文本分解为具有意义的基本文字串。

^③<http://www.sciaps.org> 1.7 A Topic Map of NIH Grants 2007, By Bruce W. Herr II, Gully Burns, David Newman, Edmund Talley Bloomington, In, 2007, Courtesy of Chalk Labs, Indiana University, Information Sciences Institute, University of California, Irvine <http://www.nihmaps.org>

图3 NIH的主题图^⑨

为了能利用PATLISYS-J文献中的人名进行各种分析,需要较为准确地合并人名,有必要采用唯一ID来识别人物。

Scopus Custom Data及其PATSTAT中同样存在着各种不同的标记(罗马字、罗马字的缩略字),区别不同标记的同一人物及标记相同的不同人物非常困难。另外,由于研究者的调动,跟踪这一变动也较困难,其结果导致了名称精度较差,没有可信性。

为此,本调查就科技文献及专利公报,开发了可以自动合并名称的系统。JST采用合并了持有的科技文献数据JSTPlus、JMEDPlus、医学和药学系数据库MEDLINE、专利公报数据(公开公报、公表公报、再公表候补)等中的著者的全部数据,把为J-Global开发的名称自动合并程序,为在Scopus Custom Data及PATSTAT中利用进行了定制。

为了确认自动名称合并的

结果,通过著者10人,进行了查全率和查准率的调查。就Scopus Custom Data而言,查全率和查准率分别平均达到了94.4%、100%,而PATSTAT,分别达到了91.7%、100%。

2.3 专利的NPL和论文信息的匹配

有关专利引文信息中的非专利文献(NPL: Non Patent Literature)和论文信息的匹配,就Scopus Custom Data的论文信息,在各个文本中包含的不同标记的变换及大写和小写的变换等基础上,作了如下的处理:

① 作为匹配的前处理,利用JST运行的科技文献检索服务体系“JDreamII”的缩略语辞书进行变换处理、删除停用词、制作了杂志名辞典。基于这些工作,可以提高匹配处理的精度。

② 抽取页信息。在信息可抽取的情况下,进行论文信息与首页和杂志名的等同匹配,在页信息无法抽取的情况下,进行首页、卷、出版年、杂志名、题目文本的全文检索匹配。

为了确保匹配结果的精度,随机抽出专利信息50件,有关与引文上NPL匹配上的论文信息,通过人工结果的核实,明确了匹配准确率大体上为100%。

2.4 共引分析

在本调查分析中,为了比较第一期、二期和三期的科学技术基本计划期间的政策地图,各自从1996年到2000年、2001年到2005年、2006年以后发行的论文中,抽取各年和各领域(生化和遗传学及分子生物学、化学工业、化学、地球科学、工程、环境科学、免疫学、材料科学、医学、交叉学科领域、神经科学、药理学和毒物学及药剂学、物理学和天文学)中被引论文数为前1%的论文,做成集合M。抽取2篇以上的这些论文,计算这些论文的相似度。作为相似度的计算方法,采用了如下的余弦函数(Salton's normalization of co-citation)^[16]。

$$K_{ij} = M_{ij} / \sqrt{M_i \cdot M_j} \quad (1.1)$$

对式(1.1) 0.3^⑩以上的论文集合,进行了聚类分析。

2.5 可视化工具

为将分析可视化、地图化,采用了将所有复杂网络可视化、可

^⑨Thomson Reuters "Research Front Methodology", <http://sciencewatch.com/about/met/rf-methodology/> (参照20100518)

解析的平台Cytoscape^①。这是开源软件项目（LGPL），由于可读入各种网络文档、基于插件的自由扩展，它成为可视化的平台。

2.6 可视化方法

作为政策地图，从PATLISYS-J系统中，可以输出以下内容：

- ① 特定领域和特定技术发展之源的业绩表示
- ② 合著关系的表示
- ③ 高被引次数论文的代表

3 分析结果

3.1 特定领域和特定技术发展之源的业绩表示

图4代表了利用关键词“Stem Cells”（干细胞），缩小数据范围，从Scopus Custom Data中抽出的论文群的引用和被引用关系的网络图。图中黄色2006年表示了2006年Cell杂志刊载的京都大学山中伸弥教授的“Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors”论文。对引用此论文的论文在何时、谁引用等事项进行了可视化。

菱形节点（◇）表示JST基金资助的研究者，非JST资助的研究者论文用方形节点（□）表示。点击某一节点，下面的Data Panel中显示论文题目、分类号、杂志名、著者所赋予的关键词、国家代码、著者、机构、是否是被专利审查官引用的论文。节点的大小表示了被引用次数的多少。

图4是以论文为对象进行的

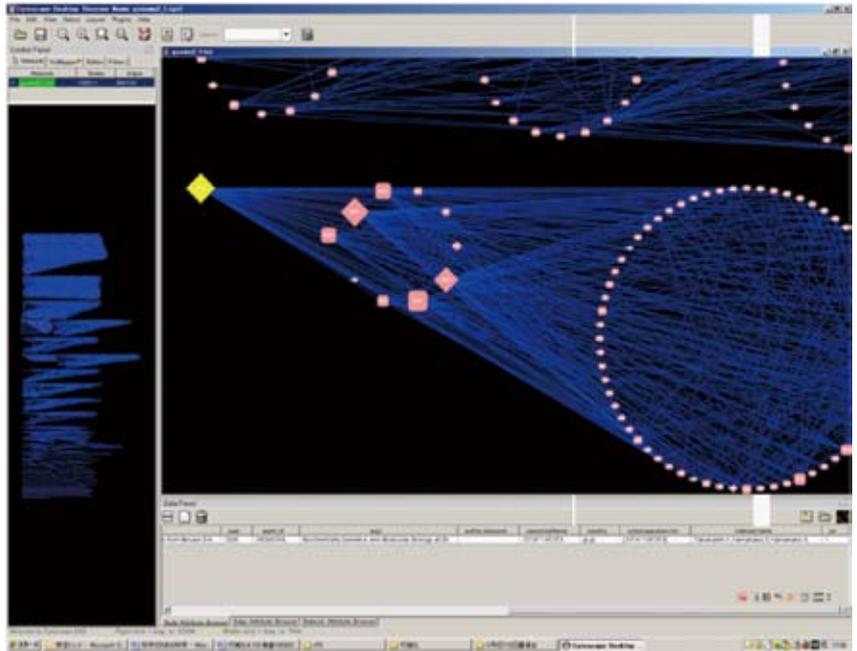


图4 Stem Cells技术的发展

分析，也可把专利作为对象进行分析。

京都大学山中伸弥教授。JST资助项目的代表研究者用菱形表示。

3.2 合著关系的表示

图5表示了上述抽出论文的全部合著关系。黄色菱形节点代表了

基于左侧垂直窗口的显示，可以纵览干细胞全部的研究者。右侧放大部分是以中山教授为中心的合著关系。

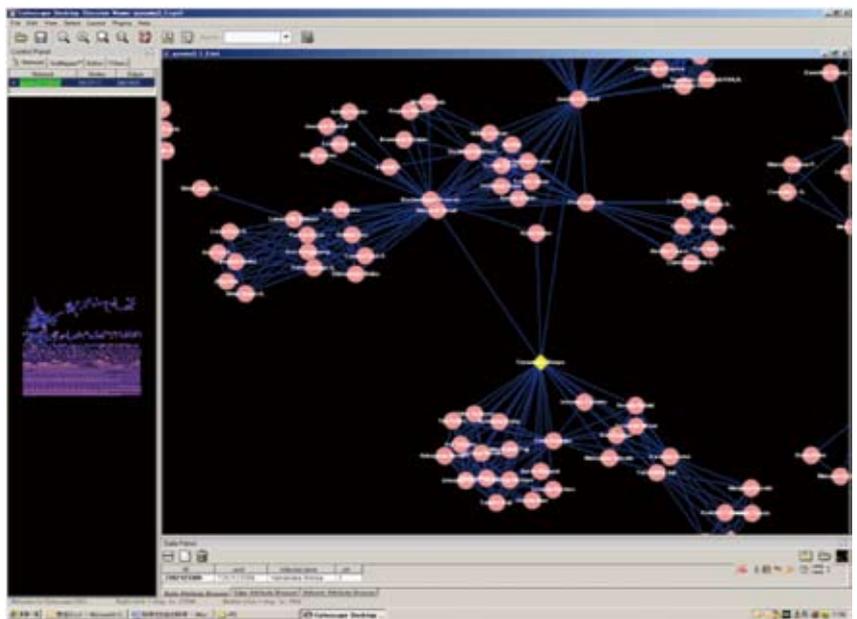


图5 合著关系的可视化

^①Cytoscape可从<http://cytoscape.org/index.php>处下载。

3.3 高被引次数论文 表示

图6左侧窗口表示了第一期科学技术基本计划期间中的世界科学技术结构。右侧是左侧窗口中特定部分的放大图。四角形为某一聚类，点击任意的四角形，在右下窗口中可以确认形成该聚类的论文题目、领域、著者赋予的关键词、著者国家、著者名、形成聚类的论文数量、这些论文是否被专利审查官引用过。

另外，菱形节点是包含JST资助研究者的论文节点。左侧图看到的红色节点为JST资助研究者的成果。构成论文的被引用次数之和代表了节点的大小。

比较图6、7、8左侧，包含JST资助研究者论文红色节点数分别为2.94%、3.86%、2.94%。分析时正值第三期科学技术基本计划中，可期待今后的增长。观察各个领域，第一期期间中化学工业，第二期期间中的免疫学领域，第三期期间中的神经科学领域红色节点较多，可确定在全体中所占的论文数量为最多。在神经科学中，与心理学相关领域、免疫学相关领域、兽医学相关领域等，表现出了横跨多个领域的特征。如图8所示，可以分析融合领域。

4 总结

基于本文的政策地图，在纵览世界动向的同时，还可检索资助领域及研究者，为此可以认为，基于

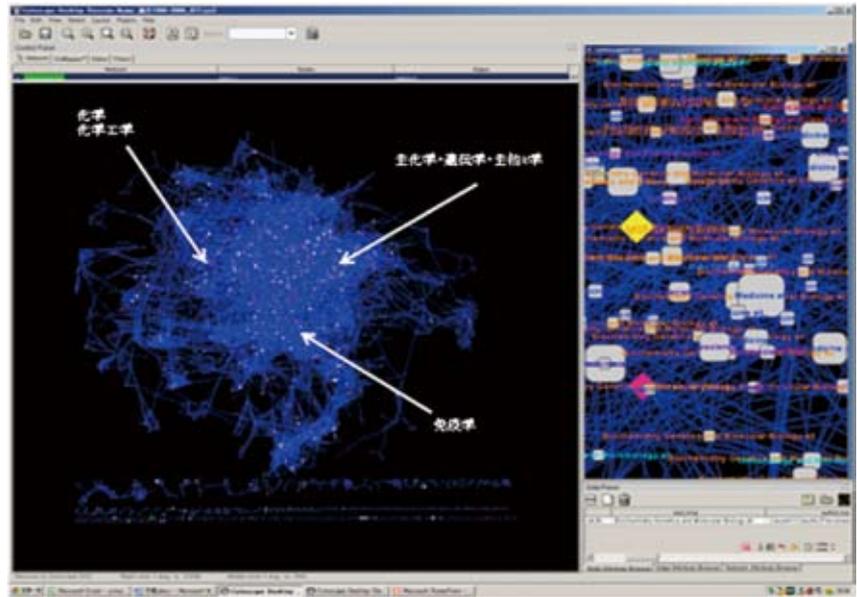


图6 第一期科学技术基本计划的政策地图

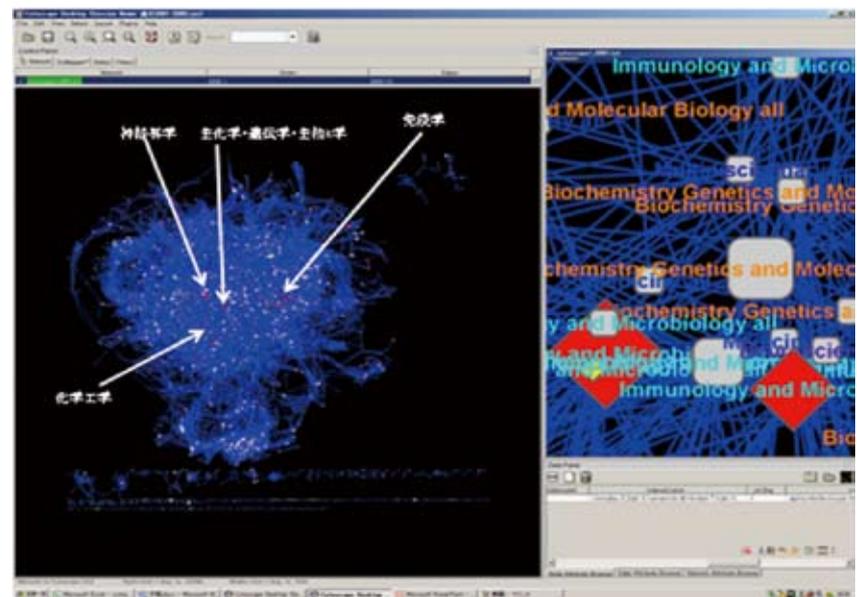
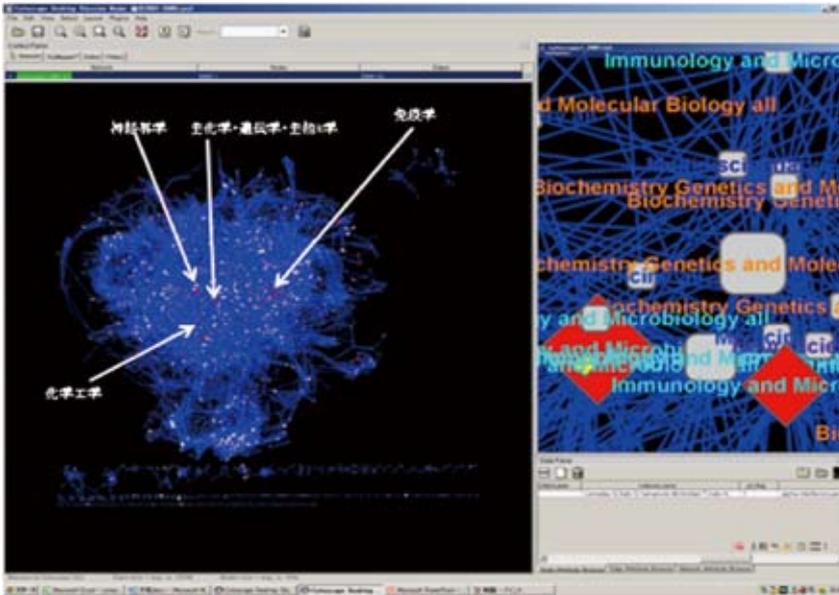


图7 第二期科学技术基本计划政策地图

科学根据 (evidence) 的战略政策的制定已接近于实现。另外，政策地图可以俯视与科学和技术密切相

关的、被专利审查官引用的论文，为此也可以探索对今后技术有影响的领域。

图8 第三期科学技术
基本计划政策地图

参考文献

- [1] 総合科学技術会議基本政策専門調査会「科学技術基本政策策定の基本方針(素案)」18ページ[EB/OL]. [2010-05-18]. <http://www8.cao.go.jp/cstp/tyousakai/seisaku/haihu07/siry02.pdf>.
- [2] 総合科学技術会議基本政策専門調査会「科学技術基本政策策定は基本方針(素案)」29ページ[EB/OL]. [2010-05-18]. <http://www8.cao.go.jp/cstp/tyousakai/seisaku/haihu07/siry02.pdf>.
- [3] PRICE D. Networks of Scientific Papers [J]. Science, 1965, 149(3683):510-515.
- [4] 財団法人新技術振興渡辺記念会. 科学技術行政政策史-その成立と発展. 科学新聞社, 2009.
- [5] Subcommittee on Social, Behavioral and Economic Sciences, committee on Science, National Science and Technology Council, Office of Science and Technology Policy, The Science of Science Policy: A Federal Research Roadmap [EB/OL]. [2010-05-18]. <http://www.scienceofsciencepolicy.net/blogs/sosp/pages/sosproadmap.aspx>.
- [6] 窪田輝藏. '科学を計る-ガフィールドとインパクト・ファクター' インターメディカル. 1996.
- [7] GARFIELD E. Research fronts [J]. Current Contents, 1994.
- [8] OECD. STI Review: Special Issue on New Science and Technology Indicators, 2002.
- [9] OECD. Science, Technology and Innovation Indicators in a Changing World: Responding to Policy Needs, 2007.
- [10] SMALL H S. Co-citation in the scientific literature: A new measure of the relationship between two documents [J]. Journal of the American society for Information Science, 1973, 24(4):265-269.
- [11] KESSLER M M, et al. Bibliographic Coupling between Scientific papers [J]. American Documentation, 1963(13):10-25.
- [12] BORNER K. Atlas of Science [M]. MIT Press, 2010.
- [13] 落合圭, 小林義英, 橋本定幸, 等. サイエンスリンクエッジによるJST事業成果分析(下) 可視化の具体的手法[J]. 情報管理, 2010,52(11):651-659.
- [14] 小林義英, 落合圭, 橋本定幸, 等. 特許の引用情報にみられる論文情報の定量的分析のためのシステム開発[J]. デジタル図書館, 2009(37):78-84.
- [15] 治部眞里, 小林義英, 落合圭, 等. サイエンスリンクエッジによるJST事業成果分析(上) 国別・機関別の分析[J]. 情報管理, 2010,52(10):601-609.
- [16] 藤垣裕子, 平川秀幸, 富澤宏之, 等. 研究評価・科学論のための科学計量学入門[M]. 丸善株式会社, 2004:70.

作者简介

治部真理 (Mari Jibu), MBA医学博士。最新研究の課題: 文献計量学 (linking)、未来学 (情景规划), E-mail: m2jibu@jst.go.jp
山崎雅和, 项目管理专业人员。最近研究の題目: 专利信息と論文情報の匹配、专利信息及論文情報のデータ清洗、基于专利和论文法整合分析世界动向。E-mail: m7yamaza@jst.go.jp

From "Science of Science" to "Science of Policy"

Mari Jibu, Masakazu Yamazaki (Original) / JST, Tokyo 102-8666, Japan
Li Ying (Translation and Editing) / Institute of Scientific and Technical Information of China, Beijing, 100038

Abstract: "Science of Science" for getting a quick overview of the structure of Science is now moving to "Science of Policy" for getting a quick overview of the structure of science with policies. On this background, it is hoped that policy making that is based on scientific evidence will be leaded and ensures transparency. This paper introduced the method and systems of visualization for the structure of science with policies and detailed analyses per researchers, with JST-PATLISYS-J.

Keywords: Science of Policy, Policy Mapping, Science Mapping Cytoscape Linkage between patents and papers

(收稿日期: 2011-02-16)