Vol.2012-ICS-165 No.1
2012/1/12

Discovering Research Themes of Institutes' Research Work

Saeed-Ul-Hassan[†1] and Ryutaro Ichise[†2]

This paper discovers the research themes of institutes' research work using analysis of scientific literature. The proposed methodology creates research profiles of the institutes by aggregating citations of highly cited works and then clusters the documents that cite those works to determine the impact, in that area of the research. Research themes are identified by clustering author defied keywords. The approach is demonstrated on several Japanese institutes in the field of Nanotechnology. The analytical techniques discussed in this paper can discover niche focus of institutes' research. This information can be very useful for the research administrators, funding agencies, and institutes leaders to understand the research structure of institutes in order to support resource allocation decisions.

1. Introduction

The quest for world-wide reputation is forcing institutions to review their performance and priorities by benchmarking their outcomes against international competitors. The global research environment has seen a marked increase in competition for financial and human resources. Moreover, science and technology continue to expand in breadth, while the cost of conducting cutting-edge research is rapidly increasing. These are some of the reasons that most countries have placed an increased emphasis on the effectiveness and efficiency of government supported research. All this means that there is a greatly increased need to be able to investigate competitive positioning in traditional and emerging niche research themes and ultimately to make more effective administrative decisions.

It is important to understand the structure of research landscape besides the strength and maturity of research in the country in order to formulate research programs effectively. One may want to consider establishing centers or tapping into existing networks of excellence depending on how research strength is distributed among the institutions in a country. For example, if one institute is obviously dominant in a particular niche research theme, then establishing a center of excellence can be an interesting option for effective use of resources. Centers of excellence can be efficient ways to utilize resources because they can provide critical mass of researchers necessary to obtain international prominence, and can function as a hub to coordinate research activity in the country, to provide capacity building, and to provide point of contact for national and international linkages.

This paper tries to discover the research themes of institutes in an important field of science i.e. Nanotechnology. At first, phenomena of Nanotechnology field diffusion into other fields of science is studied. Using this phenomena, research profiles of institutes are build which not only show institutes' focus in certain research themes, but also investigate that how these themes are diffused into other research themes. This kind of information is very useful for the research administrators, funding agencies, and institutes leaders to understand the research activities of institutes in order to support resource allocation decisions. The presented techniques in this paper are applied to available bibliography databases. Since such database resources contains the bibliographic information of research work produced by the institutes.

The present paper is organized as follows. The next section discusses relevant literature review in the context of bibliometric analysis. Section 3 describes methodology and general framework of research. Section 4 discusses experimental results. Finally, in Section 5, conclusion is discussed along with direction for future work.

†1 Asia Institute of Technology
†2 National Institute of Informatics

2. Related Work

Previously, bibliography databases have been noticeably utilized to investigate research trends and to create the structure of science.

2.1 Research Trends

Shaodong Xie and Jing Zhang[6] investigate world aerosol research trends by bibliometric analysis. They inspect the research trends of world aerosol research by looking at various perspectives including number of institutes participating each year for underlying research area. Gangfeng Zhang et al.[1] map trend of independent, national collaborative and international collaborative share of articles per year in research area of volatile organic compounds. They also identify most productive subject categories and frequency of keywords in different periods related to volatile organic compounds, however, they do not look into the strengths of individual institutes in their research work.

2.2 Maps of Science

Bibliography databases have also been utilized to create maps of science[7]. Most recently, science maps have been created to provide basic understanding of how science is organized. They are particularly useful in analyzing or comparing the disciplinary profiles of nations, funding agencies, or other institutions such as universities. For instance, disciplinary maps[5] provide a detailed template on which the funding profiles for divisions with the U.S. National Institutes of Health and National Science Foundation can be visualized and compared. Competency maps[3] allow one to identify the interdisciplinary strengths of a research, lab, institution, region, or nation. Paradigm maps[4] show the relationships between a set of related paradigms (a paradigm is the smallest possible cluster of related scientific documents). Moreover, citation analysis has been utilized to investigate the structure of science and research communities[2),8].

The above survey discusses several researches to highlight the usefulness of bibliometric analysis, to investigate the research trends and to understand the structure of science.

3. Methodology

This section describes general analysis framework of current research, Field of Research Codes (FoRs) and selected time windows for analysis.

3.1 Analysis Framework

In general, a bibliographic citation is a reference to a book, article, web page, or other published item. As the science and technology is growing rapidly, research work produced in one particular research area is being cited in many other research areas, for instance research work from Nanotechnology can be cited in general research areas such as Physical Science, Materials Engineering or Chemical Sciences or can provide the basis for future research in core Nanotechnology research area itself. This point is very interesting to investigate that the research produced by an institute in a given research area (say Nanotechnology) defuses into other research areas, or provide the basis for future research in that area. This shows the strengths of an institute in that particular research area.

Figure 1 shows general analysis framework, at first a set of papers, A, is selected in a particular research area say R1, published by a given institute, for a given time window. P:R1 shows publications in research area R1 where P:K1, P:K2, P:K3 and P:K4 are top most frequent author defined keywords from the selected set of publication. These keywords show research themes produced by an institute. Further, a set of papers, B, is selected from bibliography database; these papers cite top 1% of highly cited papers of A. Next, papers from set B are clustered using journal mapping technique using FoRs mapping (see Section 3.3). Each paper is presented in a particular conference or published in a journal and each conference or journal is classified under certain research area. Journal mapping technique uses this point and cluster the papers in respective research areas (C:R1, C:R2, C:R3, etc), where R1, R2, R3 shows the research areas. Further, percentage share of research work in research area is calculated based on number of paper each cluster contains. Finally, each cluster is further examined by the top most frequent author defined keywords (C:R3:K1, C:R3:K2, C:R3:K3, C:R3:K4). These frequent author defined keywords shows the research themes which cite the original research work of a given institute. Here C:R3:K1 can be different from

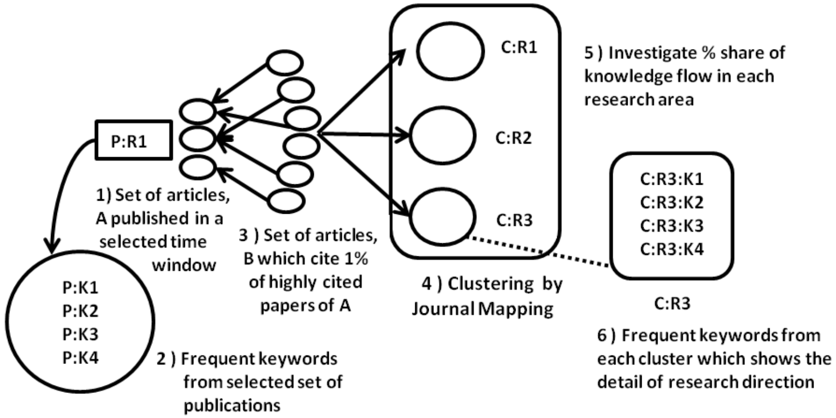

図 1 Analysis framework

P:K1.

In order to investigate the research strength of the institutes, a simple statistical technique is applied. It is done by determining outstanding institutes, which provides significant base-knowledge for the selected research area as compared to other institutes (an institute will be providing significant base-knowledge for a particular research area if the institute has great impact in that particular research area). At first, a standard model is created which shows average citation flow of all selected institutes for a particular research area, say R1. Further, an institute is considered as an outstanding institute for research area R1, which has the largest positive value of the difference between "percentage share of citation flow for R1" and "mean value of citation flow of R1 from standard model plus the standard deviation value". The rationale behind this statistics is that in normal data distribution all data points which are above mean value plus standard deviation, are extra large values.

3.2 Time Windows

Two time windows have been utilized, historical time window and recent time window. Historical time window is used to determine historical research strength of a given institute for a particular research area; however, recent time window determines the research strength of an institute in most recent years. Historical time window uses five year window for publications, e.g. publications from year 2001 to 2005, and for citations five years for each paper from the year of publication, e.g. papers published in 2004 have citation time window till year 2008. The rationale behind historical window is to give fair chance to each paper to receive citations in five years from the year of publication. However, for recent time window, publications from year 2005 to 2009 and citations received by these publications from year 2005 to 2009 have been considered. Recent time window shows recent research activity of an institute in a particular research area.

3.3 Field of Research Codes (FoRs)

Field of Research Codes (FoRs) are used to categorize research areas. The current research work utilizes FoRs to define research areas (as discussed in Section 3.1). FoRs have three levels of classification to divide subject categories. Top level FoRs contains 22 divisions including division 01 Mathematical Sciences, division 02 Physical Sciences, division 03 Chemical Sciences etc. Top level divisions are represented by two digit code and each top level division is further sub divided. Each sub division is represented by two digit code followed by the parent code (http://www.abs.gov.au/). The FoRs codes are mapped with source titles such as journals, conferences and books (http://www.era2012eidtagging.com/).

4. Experiments and Results

This section describes experimental data, experimental setting and results.

4.1 Experimental data and settings

The underlying research work utilizes Scopus database (http://www.scopus.com/), available at institute where research work was conducted (however, it is clear that provided methodology is regardless to the choice of bibliography database). Five Japanese institutes were selected to create the research profiles for the research area of Nanotechnology. At first, for each institute the articles (for both historical and recent time windows) were pulled from Scopus database, Table 1 shows statistics of experimental data. Further, the articles which cite top 1% highly cited papers of respective institute were clustered by using journal mapping technique and each cluster was assigned to the respective research area. Table 2 and 3 show citation flow of Nanotechnology using

表 1 Statistics of experimental data

	Historical Time Window		Recent Time Window	
	Publications	Citations	Publications	Citations
University of Tokyo	209	1193	228	1333
Tohoku University	105	673	159	879
AIST	86	660	201	740
NIMS	64	1054	210	1074
Osaka University	90	563	140	666

AIST = National Institute of Advanced Industrial Science and Technology,
NIMS = National Institute for Material Science

表 2 Nanotechnology citation flow (historical time window)

	University of Tokyo	Tohoku University	AIST	NIMS	Osaka University
Physical Sciences	17.3%	26.8%	15.2%	14.9%	22.9%
Nanotechnology	15.2%	11.3%	13%	15%	12.6%
Macromolecular and Materials Chemistry	9.2%	10%	14.1%	22.5%	9.2%
Chemical Sciences	5.6%	3.2%	12.1%	7.8%	13.4%
Condensed Matter Physics	5.4%	15.2%	5.4%	5.3%	7.1%
Physical Chemistry Incl. Structural	7.5%	3.9%	7.6%	8.7%	4.2%
Materials Engineering	3.2%	9.7%	4.5%	8.8%	5.3%
Analytical Chemistry	7.1%	1.5%	5.4%	2.3%	2.4%
Atomic, Molecular, Nuclear, Particle and Plasma Physics	5.6%	3%	3.40%	3.9%	3.4%
Electrical and Electronic Engineering	2.1%	3.5%	2.2%	1.7%	1.8%
Others	21.8%	11.9%	17.1%	9.1%	17.7%

AIST = National Institute of Advanced Industrial Science and Technology,
NIMS = National Institute for Material Science

Historical time window and Recent time window respectively. Research areas where citation flow is less than 1% are combined as "Others".

4.2 Results

Figure 2 shows the standard models for both historical and recent time windows (using data from Table 2 and 3). These standard models represent average percentage share of citation flow of all selected institutes. It shows overall profile of Nanotechnology research area, that how citations from Nanotechnology field flows into other areas of research. The range of citation flow in each research area (with horizontal line on each bar) represents minimum and maximum share of citation flow from selected institutes. This shows the difference in research strength of the institutes to provide impact in certain research area for future research. The results show that there is no significant

表 3 Nanotechnology citation flow (recent time window)

	University of Tokyo	Tohoku University	AIST	NIMS	Osaka University
Physical Sciences	16.2%	26.5%	15.9%	14.4%	22.6%
Nanotechnology	15.1%	13%	13.6%	15.2%	13.9%
Macromolecular and Materials Chemistry	8.9%	9.6%	13.2%	22.7%	9.6%
Chemical Sciences	6.1%	3.6%	11.7%	7.6%	13.30%
Condensed Matter Physics	5.3%	11.9%	5.5%	5.2%	5.6%
Physical Chemistry Incl. Structural	7.6%	4.7%	7.9%	8.9%	4%
Materials Engineering	3%	9.6%	4.2%	8.9%	4%
Analytical Chemistry	7.6%	1.8%	6%	2.4%	2.8%
Atomic, Molecular, Nuclear, Particle and Plasma Physics	5.3%	2.9%	3.5%	4%	4%
Electrical and Electronic Engineering	2%	3.4%	2.5%	1.7%	1.2%
Others	22.9%	13%	16%	9%	18.7%

AIST = National Institute of Advanced Industrial Science and Technology,
NIMS = National Institute for Material Science

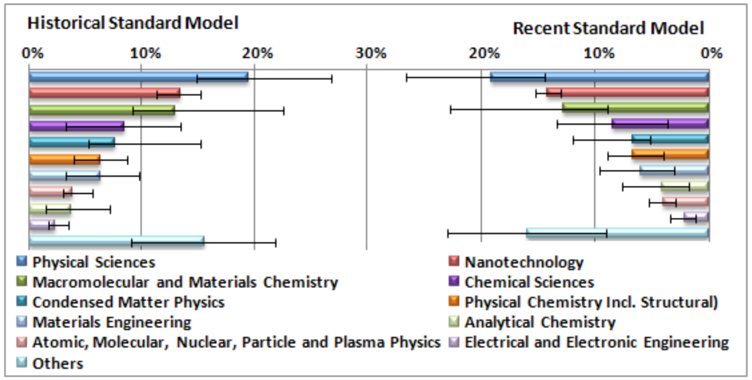

図 2 Historical and recent standard models

difference in historical and recent models; as citation flow proportion is quite same for both time windows.

4.2.1 Research Strength of Institutes

In order to investigate the research strength of the institutes in area of Nanotechnol-

表 4 Difference between the (percentage share of citation flow) and the (mean value of citation flow of recent standard model plus the standard deviation value) for each research area

	University of Tokyo	Tohoku University	AIST	NIMS	Osaka University
Physical Sciences	-0.0811	**0.021899**	-0.0841	-0.0991	-0.0171
Nanotechnology	-0.00021	-0.02121	-0.01521	**0.000793**	-0.01221
Macromolecular and Materials Chemistry	-0.09685	-0.08985	-0.05385	**0.041151**	-0.08985
Chemical Sciences	-0.06355	-0.08855	-0.00755	-0.04855	**0.008446**
Condensed Matter Physics	-0.04311	**0.022888**	-0.04111	-0.04411	-0.04011
Physical Chemistry Incl. Structural	-0.01162	-0.04062	-0.00862	**0.001383**	-0.04762
Materials Engineering	-0.06021	**0.005792**	-0.04821	-0.00121	-0.04721
Analytical Chemistry	**0.009439**	-0.04856	-0.00656	-0.04256	-0.03856
Atomic, Molecular, Nuclear, Particle and Plasma Physics	**0.004751**	-0.01925	-0.01325	-0.00825	-0.00825
Electrical and Electronic Engineering	-0.00998	**0.004015**	-0.00498	-0.01298	-0.01798

AIST = National Institute of Advanced Industrial Science and Technology,
NIMS = National Institute for Material Science

ogy, standard model with most recent time window is selected. It simply makes more sense to investigate that what is happening most recently. Table 4 shows the difference between "percent share of citation flow" and "mean value of citation flow of recent standard model plus the standard deviation value" for each research area, for all selected institutes. Results with the largest positive value are shown in bold format (see Table 4) which shows the outstanding institute for respective research areas.

As the results suggest, Tohoku University appears to be outstanding in area of Physical Science, Condensed Matter Physics, Material Engineering and Electrical and Electronically Engineering. Whereas, National Institute for Material Science appears to be outstanding in areas of Macromolecular and Materials Chemistry, Physical Chemistry Incl. Structural and Nanotechnology. Meanwhile, publications produced by University of Tokyo in area of Nanotechnology significantly cite in Analytical Chemistry and Atomic Molecular Nuclear Particles and Plasma Physics. It is noted that the University of Tokyo and National Institute for Material Science are quite close as the outstanding institutes to provide the basis of research in Nanotechnology and these institutes are good to be considered as the centers of excellence in research area of Nanotechnology.

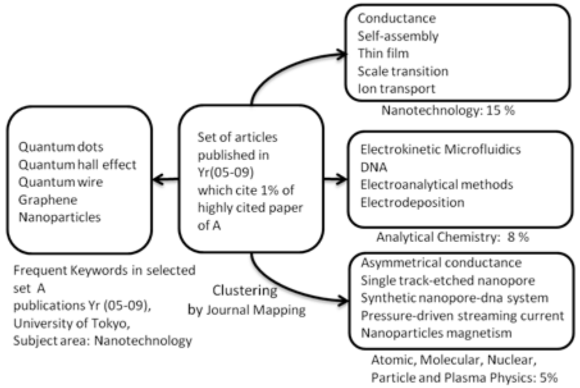

図 3 Nanotechnology research profile of University of Tokyo

4.2.2 Research Themes Institutes

Figure 3 shows research profile of University of Tokyo in Nanotechnology. Quantum dots, Quantum hall effect, Quantum wire, Graphene and Nanoparticles (left side of Figure 3) are the top most frequent keywords (author defined keywords) in the research work produced by University of Tokyo. Moreover, Figure 3 also shows some of the prominent research areas which cite research work of University of Tokyo along with the most frequent keywords (right side of Figure 3). It shows that what kind of research work is carried out in each respective research area, e.g. research work in Nanotechnology it self is utilized in the concepts such as Conductance, Self-assembly, Thin films, Scale transition and Ion transport.

Figure 4 shows research profile of National Institute for Material Science. Nanoparticles, Self-assembly, Quantum dots, Mesoporous silica, and Photoluminescence (left side of Figure 4) are the top most frequent research themes within Nanotechnology produced by the institute. Moreover, Figure 4 also shows some of the prominent research areas where Nanotechnology diffuses along with the most frequent themes in each research area (right side of Figure 4), for instance, research diffuses in Macromolecular and Materials Chemistry and is utilized in themes such as Mesoporous, Mesoporous silica, Ionic liquid, and B1. nanomaterials.

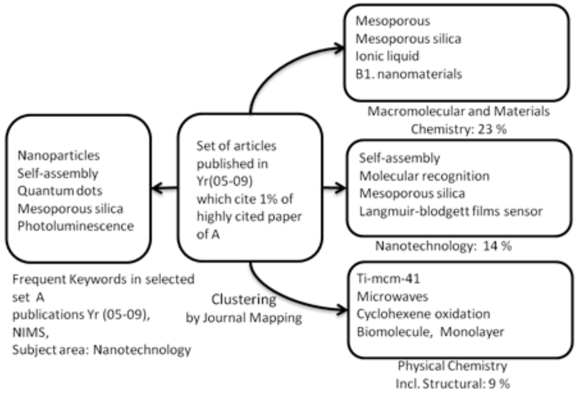

図 4 Nanotechnology research profile of National Institute of Material Science

The analysis discovers that research work conducted by University of Tokyo focuses on themes such as Quantum hall effect, Quantum wire and Graphene. However, National Institute for Material Science shows different focus as compare to University of Tokyo i.e. Self-assembly, Mesoporous silica and Photoluminescence. We also see that research themes of both institutes do not diffuse into same research areas with same proportion. For instance, research work in Nanotechnology produced by University of Tokyo diffuses into Nanotechnology, Analytical Chemistry and Atomic, Molecular, Nuclear, Particle and Plasma Physics with 15%, 8% and 5% percentage share of research work respectively. However, research work in Nanotechnology produced by National Institute for Material Science diffuses into Macromolecular and Materials Chemistry, Nanotechnology and Physical Chemistry Incl. Structural with 23%, 14% and 9% percentage share of research work respectively. We see that research work produced by the institutes can diffuse into different research areas and research themes, which shows the nature of the research conducted at the institutes.

5. Conclusion

In this paper we have discovered research themes of institutes' research work in the field of Nanotechnology using analysis of scientific literature. We have shown that how Nanotechnology field is diffused into other areas of research. We have also investigated research themes around the fields where Nanotechnology field diffuses. Finally, we illustrate research profiles of University of Tokyo and National Institute for Material Science.

The presented analysis provide the insight of research work conducted at the institute. This kind of information is very useful for the research administrators, funding agencies, and institutes leaders to knit together various niche strengths of institutes and to understand the research activities of institutes in order to support resource allocation decisions. In future, we plan to create research profiles of institutes for various disciplinary and inter-disciplinary research areas.

参 考 文 献

1) Gangfeng, Z., Shaodong, X., Yuh-Shan, H.: A bibliometric analysis of world volatile organic compounds research trends. Scientometrics 83(2), 477–492 (2010)
2) Ichise, R., Takeda, H., Ueyama, K.: Community mining tool using bibliography data. In: Proceedings of the 9th International Conference Information Visualization. pp. 953–958. IEEE Computer Society, Washington, DC, USA (2005)
3) Kevin, B., Katy, B., Richard, K.: Mapping the structure and evolution of chemistry research. In: Proceedings of the 11 th International Conference of Scientometrics and Informetrics. pp. 112–123 (2007)
4) Richard, K., Kevin, B.: Identifying a better measure of relatedness for mapping science. Journal of the American Society for Information Science and Technology 57(2), 251–263 (2006)
5) Richard, K., Kevin, B.: Toward a consensus map of science. Journal of the American Society for Information Science and Technology 60(2), 455–476 (2009)
6) Shaodong, X., Jing, Z.: Assessment of world aerosol research trends by bibliometric analysis. Scientometrics 77(1), 113–130 (2008)
7) Small, H.: Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for Information Science 24(4), 265–269 (1973)
8) Small, H., Phineas, U.: Citation structure of an emerging research area on the verge of application. Scientometrics 79(2), 365–375 (2009)