Global nanotechnology research literature overview†

*Ronald N. Kostoff *, Raymond G. Koytcheff and Clifford G. Y. Lau*

Text mining was used to extract technical intelligence from the open source global nanotechnology and nanoscience research literature (SCI/SSCI databases). The following were identified: (i) the nanotechnology/nanoscience research literature infrastructure (prolific authors, key journals/institutions/countries, most cited authors/journals/documents); (ii) the technical structure (pervasive technical thrusts and their inter-relationships); (iii) nanotechnology instruments and their relationships; (iv) potential nanotechnology applications; (v) potential health impacts and applications, and (vi) seminal nanotechnology literature. The results are summarized in this article.

Keywords: Bibliometrics, document clustering, nanoparticle, nanotechnology, nanotube, text mining.

NANOTECHNOLOGY is booming! In the global fundamental nanotechnology research literature as represented by the Science Citation Index/Social Science Citation Index (SCI/SSCI)[1], global nanotechnology publications have grown dramatically in the last two decades.

Due to this exponential growth of the global nanotechnology open literature, there is a need for gaining an integrated quantitative perspective on the state of this literature. In 2003–05, a comprehensive text-mining study was performed to survey the technical structure and infrastructure of the global nanotechnology research literature, as well as the seminal nanotechnology literature[2,3]. Based on the wide-scale interest generated by these reports, it was decided to update and expand the study using more recent data, a much more comprehensive query and more sophisticated analytical tools.

In the updated study, text mining was used to extract technical intelligence from the open source global nanotechnology and nanoscience research literature (SCI/SSCI databases). The following were identified: (i) the nanotechnology/nanoscience research literature infrastructure (prolific authors, key journals/institutions/countries, most cited authors/journals/documents); (ii) the technical structure (pervasive technical thrusts and their inter-relationships); (iii) nanotechnology instruments and their relationships; (iv) potential nanotechnology applications; (v) potential health impacts and applications, and (vi) seminal nanotechnology literature. The results are summarized in this article. A more detailed report on the results and methodologies of this updated study can be found in Kostoff *et al.*[4].

This article is an overview of the highlights of the total study, including the production efficiency of seminal nanotechnology documents. The results are divided into four main sections: Infrastructure, Technical structure, Instrumentation and Applications. The Applications section is further divided into non-medical and medical. The results will be presented in the order listed above. Next, the seminal nanotechnology literature production efficiency will be presented.

Infrastructure describes the performers of nanoscience/nanotechnology research at different levels, ranging from individual to national performers, and it includes archived literature as well. Technical structure identifies the pervasive technical thrusts (and their inter-relationships) of the nanoscience/nanotechnology literature. Instrumentation provides both infrastructure and technical structure of the subset of the nanoscience/nanotechnology literature that addresses specific instruments. Applications provides the infrastructure and taxonomy of the subset of the nanoscience/nanotechnology literature that addresses specific non-medical and medical applications.

Approach

An extensive nanotechnology/nanoscience-focused query (300 + terms) was applied to the SCI/SSCI database. The nanotechnology/nanoscience research literature technical structure (taxonomy) was obtained using computational linguistics, especially document clustering. The nanotechnology/nanoscience research literature infrastructure (prolific authors, key journals/institutions/countries, most

†The views expressed in this article are solely those of the authors, and do not represent those of the Department of the Navy or any of its components, or the Institute for Defense Analyses.

Ronald N. Kostoff and Raymond G. Koytcheff are in the Office of Naval Research, 875 N. Randolph Street, Arlington, VA 22217, USA and Clifford G. Y. Lau is in the Institute for Defense Analyses, 4850 Mark Center Drive, Alexandria, VA 22311, USA.
*For correspondence. (e-mail: kostofr@onr.navy.mil)

cited authors/journals/documents) for each of the clusters generated by the document clustering algorithm was obtained using bibliometrics.

The instrumentation literature associated with nanoscience and nanotechnology research was examined. About 65,000 nanotechnology records for 2005 were retrieved from the SCI/SSCI, and ~27,000 of these were identified as instrumentation-related. All the diverse instruments were identified and their associated documents categorized in a hierarchical taxonomy. Metrics associated with research literature for specific instruments/instrument groups were generated.

The applications literature associated with nanoscience and nanotechnology research was examined. Through visual inspection of 60,000 of the abstract phrases of the same downloaded 2005 records, all the diverse non-medical applications were identified and their associated documents categorized in a hierarchical taxonomy. Metrics associated with research literature for specific applications/applications groups were generated.

For medical applications, a fuzzy clustering algorithm (where a record could be assigned to multiple clusters) was applied to the downloaded 2005 records. A sub-network that encompassed all the medical applications was identified. Again, metrics associated with research literature for specific medical applications were generated.

Results

Infrastructure

Country publications:

- Global nanotechnology research article production exhibited exponential growth for more than a decade (Figure 1).
- The most rapid growth over that time period came from East Asian nations, notably China and South Korea (Figure 2).

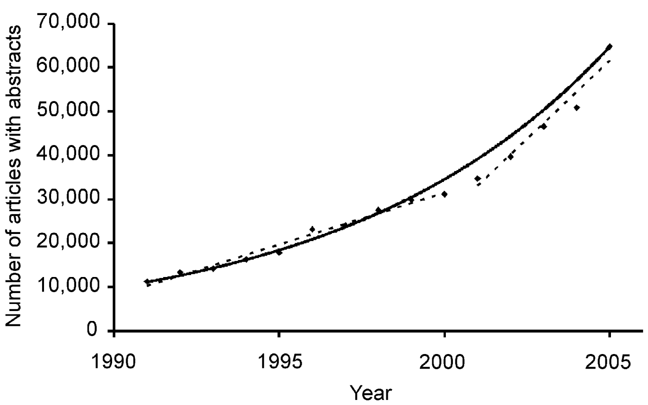

Figure 1. SCI/SSCI articles vs time: total records retrieved.

- Some of this apparent rapid growth (in China, for example) is partially due to (i) a country's researchers publishing a non-negligible fraction of total papers in domestic low impact factor journals, and (ii) these journals being accessed recently by the SCI/SSCI, rather than due to growth based on increased sponsorship or productivity.
- China's representation in high impact factor journals was small, but increasing.
- From 1998 to 2002, China's ratio of high impact nanotechnology papers to total nanotechnology papers doubled, placing the country at parity for this metric with the advanced nations of Japan, Italy and Spain.
- The US remained the leader in aggregate nanotechnology research article production.
- In some selected nanotechnology sub-areas, China had achieved parity or taken the lead (see Figure 3 for the nanocomposites example).

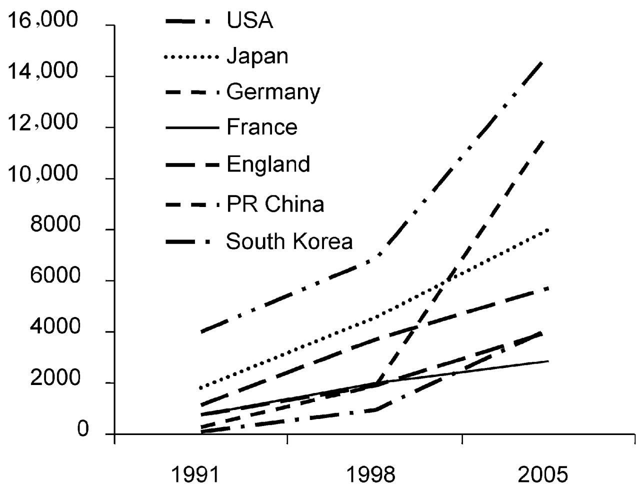

Figure 2. Country comparison time trend (number of articles vs time).

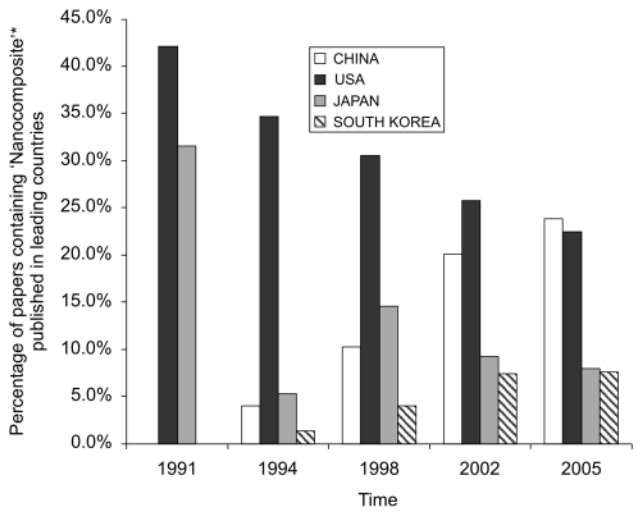

Figure 3. Number of papers containing 'nanocomposite'*.

- South Korea started even further behind China in both total nanotechnology publications and highly cited papers, but has advanced rapidly to become a second-tier contender in total and highly cited papers.

Country citations:

- There was a clear distinction between the publication practices of the three most prolific Western nations (USA, Germany, UK) and the three most prolific East Asian nations (China, Japan, South Korea). The Western nations published in journals with almost twice the weighted average impact factors of the East Asian nations. Much of the difference stems from the East Asian nations publishing a non-negligible amount in domestic low impact factor journals, while the Western nations publish in higher impact factor international journals.
- Two countries that led in production of the most cited nanotechnology papers were the US (126) and Germany (31). They accounted for 40% of the most cited nanotechnology papers.
- The high paper volume production East Asian countries of China and South Korea accounted for 2% of the most cited nanotechnology papers.
- Despite the increased paper productivity from East Asian countries, the US continued to generate the most cited nanotechnology papers.

Technical structure

The total retrieved nanotechnology database for 2005 was examined from four perspectives to identify pervasive thematic thrusts: document clustering, autocorrelation mapping, factor analysis and cross-correlation mapping. Each perspective provided valuable insights on the fundamental nanotechnology literature structure. Only document clustering results are presented here.

Document clustering: The database was divided into 256 thematic clusters by the clustering algorithm. USA produced most papers in 169 thrusts, China led in 70, Japan led in 15, and India, South Korea and Spain each led in one.

A hierarchical taxonomy was constructed from these 256 elemental clusters. Of the sixteen fourth-level categories in taxonomy, China was the publication leader in six. Specifically, China led in: Properties of thin films; Diamond films; Applications of carbon nanotubes; Multi-walled nanotubes; Nanomaterials and nanoparticles, and Polymers, composites and metal complexes (Figure 4; categories with solid shading denote publication lead by China, and those with vertical lines and shading denote publication lead by Japan. Light shading means category leader has 100–125% of the USA publications; medium shading 125–150%; dark shading >150%). Essentially, China led in the materials and nanostructures component of the database, whereas USA led in the physical science phenomena and biomedical components.

Instrumentation

A wide variety of instruments are used in nanoscience and nanotechnology research. Key among these are X-ray diffraction (XRD), electron microscope variants, atomic force microscopy, scanning tunnelling microscopy and spectroscopy variants.

LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4
Quantum phenomena, Optics, Electronics, Magnetism, Tribology, and Films (32,983 records)	Quantum phenomena, Optics, Electronics, Magnetism, and Tribology (26,077 records)	Quantum phenomena (3326 records)	Quantum dots (2028 records)
			Quantum wells, Wires, and States (1298 records)
		Optics, Electronics, Magnetism, and Tribology (22,751 records)	Optics and Electronics (16,432 records)
			Magnetism and Tribology (6319 records)
	Films (6906 records)	Thin films (4760 records)	Properties of thin films (2251 records)
			Applications of thin films (2509 records)
		Deposition of films (2146 records)	Deposition of thin films (1752 records)
			Diamond films (394 records)
Nanotubes, Nanomaterials, Nanoparticles, Polymers, Composites, Metal complexes, and Bionanotechnology (31,742 records)	Nanotubes (3211 records)	Multi-walled nanotubes (2350 records)	Applications of carbon nanotubes (474 records)
			Multi-walled nanotubes (1876 records)
		Single-walled nanotubes (861 records)	Single- and double-walled nanotubes (447 records)
			Single-walled nanotubes (414 records)
	Nanomaterials, Nanoparticles, Polymers, Composites, Metal complexes, and Bionanotechnology (28,531 records)	Nanomaterials, Nanoparticles, Polymers, Composites, and Metal complexes (22,686 records)	Nanomaterials and Nanoparticles (14,263 records)
			Polymers, Composites, and Metal Complexes (8423 records)
		Bionanotechnology (5845 records)	DNA (775 records)
			Proteins and Cellular components (5070 records)

Figure 4. Four-level hierarchical taxonomy.

AFM, NMR, Calorimetry (8423)	NMR, RS, Calorimetry (4684)	NMR, Complexes, Compounds (1546)	NMR, Spectroscopy (306)
			NMR, Complexes, Compounds (1240)
		RS, Calorimetry (3138)	DSC (1138)
			Raman scattering, RS, AFM (2000)
	AFM (3739)	AFM, Films, Tip, Imaging (2003)	AFM, Film, Tip, Imaging (1055)
			AFM, Film, Substrate, Deposit (948)
		AFM, Films, Deposition, Growth, Substrate (1736)	AFM, Film, Deposit, Substrate, Growth (1511)
			AFM, Magnetic (226)
EM, XRD (19,090)	EM (4492)	TEM (2545)	HRTEM (296)
			TEM (2249)
		SEM, Films, Composites, Particles, Cells (1947)	SEM, Film, Particle, Cell (1652)
			SEM, IS (295)
	XRD, Films (14,598)	SEM, XRD, Films, Coatings, Composites (3634)	SEM, XRD (1451)
			SEM, Film, Coating, Deposit, XRD (2183)
		XRD, TEM, Thin films (10,964)	TEM, Film, Particle, Nanoparticle, STM (5986)
			Film, XRD, XPS (4978)

Figure 5. Nanotechnology instrumentation taxonomy. AFM, Atomic Force Microscopy; NMR, Nuclear Magnetic Resonance; EM, Electron Microscopy; XRD, X-Ray Diffraction; RS, Raman Spectroscopy; TEM, Transmission Electron Microscopy; HRTEM, High Resolution Transmission Electron Microscopy; SEM, Scanning Electron Microscopy; DSC, Differential Scanning Calorimetry; IS, Infrared Spectroscopy, and STM, Scanning Tunnelling Microscopy.

Instrument taxonomy: Hierarchical taxonomy offered the following insights:

- In this nanotechnology instrumentation study, China produced about 25% more papers than the USA (Figure 5; shading represents China's publication leadership; darker shading represents stronger publication leadership). By contrast, in the full nanotechnology study, USA produced about 25% more papers than China.
- Much of China's over-production occurred in the XRD-related categories, but there was some over-production in transmission electron microscopy and NMR and calorimetry-related categories as well.
- The US dominance was in atomic force microscopy.
- Because of the large Chinese and South Korean contributions to the nanotechnology instrumentation literature, author-name analysis at aggregate levels was not effective; Asian names are usually monosyllable, many times with no middle names. Due to the relatively high frequency of paper publications, there is good possibility that the same last name represents multiple authors. Potential name disambiguation is under study.
- Even though USA has a large presence overall, relatively few US institutions were listed among the most prolific in the nanotechnology instrumentation papers. The Asian and European efforts appeared concentrated in relatively few but large institutions.

Applications

The study also identified the main nanotechnology applications, both medical and non-medical, as well as the related science and infrastructure. These relationships will allow the potential user-communities to become involved with the applications-related science and performers at the earliest stages, to help guide the science conversion towards specific user needs most efficiently.

Non-medical applications: Applications thrust areas – Factor analysis.

Factor analyses were performed to show the thematic areas in non-medical applications. A six-factor analysis showed the following themes:

- Factor 1: Optoelectronics
- Factor 2: Tribology
- Factor 3: Lithography
- Factor 4: Control systems
- Factor 5: Devices
- Factor 6: Microsystems.

Applications thrust areas – Factor analysis and visual inspection.

The main non-medical applications thrust areas identified above were augmented by important but non-networked thrusts, and the nine resulting themes were related to science and infrastructure by co-occurrence matrices. Also, the total non-medical applications was combined into one unit, and related to science and infrastructure by co-occurrence matrices. For non-medical applications:

- USA led in total non-medical applications publications and in six out of nine themes in high-tech research areas such as devices, sensors and lithography.

China led in publications in three traditional areas: catalysis, tribology and electrochemistry.
- In total non-medical applications, two of the top three institutions were Chinese. However, USA was well represented by the large State University systems of the University of California and University of Illinois.
- The journal *Applied Physics Letters* appeared in the top layer in seven of the nine themes and was by far the leader in total non-medical applications publications. *Journal of Physical Chemistry B* appeared in four of the nine themes, as also *Journal of Applied Physics*.

Medical applications: Applications thrust areas – Visual inspection/fuzzy clustering.

A medical applications categorization constructed from visual inspection of the detailed fuzzy clustering categories showed five broad thematic categories:

- Cancer treatment
- Sensing and detection
- Cells
- Proteins
- DNA.

Applications thrust areas – Fuzzy clustering.

For medical applications, analysis of nineteen thematic categories obtained from fuzzy clustering of the total 2005 nanotechnology database revealed the following:

- USA was the publication leader in total health types, and in all the thematic areas as well, mostly by a wide margin. China was the second most prolific in seven thematic areas, Japan in six, Germany in four and England in two.
- The University of California system led in five clusters, the Chinese Academy of Science led in four, and the National University of Singapore led in three. The University of California and the Chinese Academy of Science were the most prolific in the non-medical applications as well, but their orders were reversed. The National University of Singapore was a prolific contributor, especially in pharmaceuticals and biomaterials.
- The journal *Langmuir* contained the most nanotechnology articles in total health, and was in the top layer of ten of nineteen themes. The only journals in common in the top layers of applications and health were *Langmuir* and *Journal of Physical Chemistry B*.

Production efficiency of global nanotechnology literature

The global nanotechnology research literature has two main components: spatial and temporal. The spatial component covers present-day nanotechnology research being conducted globally. The temporal component reflects the impact that vintage literature has had on modern-day nanotechnology research.

Both the temporal and spatial components need to be understood for full comprehension of global nanotechnology research, and for the establishment of strategic nanotechnology policy. Assessment tools and processes have advanced sufficiently to allow an integrated picture of nanotechnology to be obtained.

The summary material presented earlier concentrates on the spatial component. The remainder of this article will concentrate on one aspect of temporal component, production efficiency of the seminal nanotechnology literature.

All the nanotechnology documents published between 1991 and 2005 were downloaded. Then, the subset with the highest number of citations was extracted, and a text mining analysis of that subset was performed to obtain the characteristics of the most cited nanotechnology documents[4]. Following this, the relationship between document production and seminal paper production for countries was identified.

Relation of seminal nanotechnology document production to total nanotechnology document production: There is a substantial value in understanding the efficiency of seminal nanotechnology document production, i.e. the ratio of seminal nanotechnology documents produced to over-all nanotechnology documents produced. The present short section addresses some methods for arriving at this ratio.

Citations (and publications) for nanotechnology documents published in two specific years were examined. The purpose was to obtain some time trend data as well as better statistics than one year's data could provide. All nanotechnology documents for 1998 and 2002 were retrieved and analysed. These years were selected to be as close to the present as possible, in order to insure currency of findings, yet sufficiently vintaged to insure accumulation of adequate citations.

Normalized country production of seminal nanotechnology papers: The main nanotechnology query in this study[4] was used to retrieve documents from the SCI/SSCI for 1998 and 2002. Distribution of number of publications among institutions and countries was generated using the Analyze function of the SCI search engine. Then, the publications for each year were ordered according to Time cited. The most highly cited publications were extracted, and the country and institution distributions for those documents were generated. The country and institution publication distributions were then compared to the citation distributions. This allowed identification of countries whose citation fractions were greater than their publication fractions (and thus were producing highly cited papers more efficiently than their publication statistics would predict).

A central issue was how one defines most highly cited papers. Are these seminal papers the top 10, top 100 or top 1%? Because of the discrete choice imposed by the Analyze function at present, results for the top 100, 250 and 500 documents were examined parametrically. While some re-ordering occurred, countries producing seminal documents were plainly evident at the top of the list. Therefore, results using the 500 most cited documents (about 1% of the total documents retrieved for 2002, and about 1.5% of the total documents retrieved for 1998) are presented.

Table 1 shows the country distributions for 1998. The left column in Table 1 shows ranking according to a country's total nanotechnology publications in 1998. For example, in 1998 USA produced 25.99% of the total nanotechnology publications. The right column in Table 1 shows ranking according to a country's representation on most highly cited papers. For example, USA was represented on 58.8% of the 500 most highly cited nanotechnology papers published in 1998.

Table 1. Country distributions – Overall records/500 most cited records (1998)

Country rank by total publications		Country rank by most cited records (121 cites min)	
Country	Percentage	Country	Percentage
USA	25.99	USA	58.80
Japan	15.72	Germany	12.20
Germany	13.72	Japan	9.60
France	7.73	France	8.00
England	6.93	England	7.80
P. R. China	6.10	Switzerland	4.20
Russia	4.87	The Netherlands	3.20
Italy	3.89	Canada	2.40
Spain	3.02	Israel	2.40
South Korea	2.96	Italy	2.20
Canada	2.81	Sweden	1.80
Switzerland	2.44	Spain	1.60
India	2.31	Australia	1.40
Sweden	2.13	P. R. China	1.40
The Netherlands	1.88	Austria	1.20
Poland	1.68	India	1.00
Taiwan	1.63	Russia	1.00
Australia	1.52	Denmark	0.80
Belgium	1.32	Ireland	0.80
Israel	1.27	Belgium	0.60
Brazil	1.20	Brazil	0.40
Denmark	0.94	Finland	0.40
Austria	0.89	Hong Kong	0.40
Ukraine	0.78	Hungary	0.40
Scotland	0.76	Scotland	0.40
Mexico	0.71	South Korea	0.40
Czech Republic	0.69	Croatia	0.20
Finland	0.67	Czech Republic	0.20
Hong Kong	0.66	North Ireland	0.20
Hungary	0.65	Norway	0.20
Singapore	0.65	Poland	0.20

Thus USA was both the most prolific nanotechnology publishing country and most represented country on highly cited nanotechnology papers for 1998. Its ratio of per cent representation on most highly cited nanotechnology papers to per cent of total nanotechnology publications (ratio = 58.80/25.99) was 2.26. A ratio greater than one indicates that a country has higher representation on most cited papers than would be expected from its publications alone. A ratio less than one indicates that a country has lower representation. A ratio of 2.26 for USA indicates that the country's representation on most highly cited records is 2.26 times what would be expected based on nanotechnology publications alone.

None of the other producers has ratios approaching that of USA (for 1998 publications), and only some of the smaller hi-tech countries (Switzerland, the Netherlands, Israel) had ratios that only remotely approach that of USA.

Table 2. Country distributions – Overall records/500 most cited records (2002)

Country rank by total publications		Country rank by most cited (80 cites min)	
Country	Percentage	Country	Percentage
USA	24.02	USA	58.20
Japan	15.09	Germany	11.40
P. R. China	11.62	Japan	8.40
Germany	11.55	England	6.20
France	7.43	P. R. China	5.80
England	5.86	France	5.40
Russia	4.83	South Korea	3.80
South Korea	4.45	Switzerland	3.40
Italy	3.92	Canada	2.80
Spain	3.09	The Netherlands	2.20
India	2.89	Italy	2.00
Canada	2.40	Spain	2.00
Taiwan	2.18	Sweden	2.00
Sweden	2.05	Finland	1.40
Poland	1.92	Belgium	1.20
Brazil	1.91	Brazil	1.20
Switzerland	1.80	Denmark	1.20
The Netherlands	1.77	Russia	1.20
Australia	1.54	Australia	1.00
Belgium	1.26	Austria	1.00
Israel	1.25	Israel	1.00
Singapore	1.22	Scotland	0.80
Austria	1.02	Singapore	0.80
Ukraine	0.99	Taiwan	0.60
Mexico	0.81	India	0.40
Scotland	0.78	Ireland	0.40
Czech Republic	0.78	Portugal	0.40
Finland	0.73	Argentina	0.20
Denmark	0.69	Czech Republic	0.20
Portugal	0.62	Greece	0.20
Hungary	0.59	Hungary	0.20
Greece	0.56	Lithuania	0.20
Turkey	0.51	Mexico	0.20
Argentina	0.46	Poland	0.20
Romania	0.45	Slovenia	0.20
Bulgaria	0.31	Turkey	0.20

Countries that have exhibited rapid growth in SCI/SSCI nanotechnology paper production in recent years (e.g. China, South Korea) have ratios an order of magnitude less than that of USA (for 1998).

Table 2 shows the same type and structure of data as Table 1, but for 2002. The USA remains dominant in nanotechnology publications and representation on most highly cited nanotechnology papers, with a ratio of 2.42. A few of the smaller Central/Northern European countries (Switzerland, Finland, Denmark) have ratios on the order of two, and form the second ratio tier after the USA. Norway, the third member of the small Scandanavian countries, has about 1/3 the publications of Finland/ Denmark, and has no representation on the 500 most cited papers list, in line with its relatively poor citation performance shown in our Finland country assessment study[5].

A number of countries retain the same ratio as in 1998 (within 10%), including the USA, Germany, Japan, England, Switzerland, Italy and Spain. China's ratio doubled to about 0.5, placing it on parity with Japan, Italy and Spain for this metric. In a recent study by the first author[6], it was shown that China's growth of papers in high impact factor journals was faster than its rate of overall publication growth, and that conclusion may be reflecting itself in the present numbers. South Korea's ratio jumped even more dramatically from 1998. Russia's, Taiwan's and Poland's ratios remain low, and India's ratio decreased substantially to join this latter group for 2002.

1. SCI, 2006, Certain data included herein are derived from the Science Citation Index/Social Science Citation Index prepared by the THOMSON SCIENTIFIC® Inc. (Thomson®), Philadelphia, Pennsylvania, USA; ©Copyright THOMSON SCIENTIFIC® 2006. All rights reserved.
2. Kostoff, R. N., Stump, J. A., Johnson, D., Murday, J. S., Lau, C. G. Y. and Tolles, W. M., The structure and infrastructure of the global nanotechnology literature. *J. Nanopart. Res.*, 2006, **8**, 301–321.
3. Kostoff, R. N., Murday, J. S., Lau, C. G. Y. and Tolles, W. M., The seminal literature of global nanotechnology research. *J. Nanopart. Res.*, 2006, **8**, 193–213.
4. Kostoff, R. N., Koytcheff, R. G. and Lau, C. G. Y., Structure of the global nanoscience and nanotechnology research literature. DTIC Technical Report (http://www.dtic.mil/). Defense Technical Information Center, Fort Belvoir, VA, USA, 2007.
5. Kostoff, R. N., Tshiteya, R., Bowles, C. A. and Tuunanen, T., The structure and infrastructure of the finish research literature. *Technol. Anal. Strategic Manage*., 2006, **18**, 187–220.
6. Kostoff, R. N., Briggs, M. B., Rushenberg, R., Bowles, C. and Pecht, M., The structure and infrastructure of Chinese science and technology. DTIC Technical Report (http://www.dtic.mil/). Defense Technical Information Center, Fort Belvoir, VA, USA, 2006.

Received 1 February 2007; revised accepted 16 February 2007