E-Commerce Research Trend Forecasting: A Study of Bibliometric Methodology

Hsu-Hao Tsai*, Johannes K. Chiang
Department of Management Information System, National Chengchi University
98356512@nccu.edu.tw*, jkchiang@nccu.edu.tw
doi:10.4156/jdcta.vol5. issue1.12

Abstract

This p aper surve ys e-com merce (EC) tech nology t rends and forecasts u sing bibliometric analysis from 1 989 to 200 9 w ith topi c as "e-com merce" i n S SCI da tabase. T he bi bliometric analytical technique is used to examine the topic in SSCI journals from 1989 to 2009, based on the scope of 2655 literatures of EC. This paper implemented and classified EC literatures using the seven catego ries f or diff erent distribution statu s in orde r t o explore h ow EC t echnology trends and applications have developed in th is period. Also, the pap er will perform K-S test to verify the reliability of Lotka's Law. The analysis provides a roadmap to guide future research and a bstract the t rend inf ormation so that EC researchers c an save some t ime sin ce c ore knowledge will be concentrated in core categories. This implies that the phenomenon "success breeds success" is more common in higher quality publications.

Keywords*: E-commerce, Technology Trend and Forecast, Bibliometric Methodology*

1. Introduction

Electronic commerce, commonly known as (electronic marketing) e-commerce or ecommerce (EC), consists of the buying and selling of products or services over electronic systems such as the internet and other related networks. The amount of trade conducted electronically has grown extraordinarily with widespread Internet usage. The use of commerce is conducted in this way, spurring and drawing on innovations in electronic funds transfer, supply chain management, Internet marketing, online transaction processing, electronic data interchange (EDI), inventory management systems, and automated data collection systems. Modern electronic commerce typically uses the World Wide Web at least at some point in the transaction's lifecycle, although it can encompass a wider range of technologies such as e-mail as well.

A large percentage of EC is conducted entirely electronically for virtual items such as access to premium content on a website, but most EC involves the transportation of physical items in some way. Online retailers are sometimes known as e-tailers and online retail is sometimes known as e-tail. Almost all big retailers have EC presence on the world wide web.

EC that is conducted between businesses is referred to as business-to-business or B2B. B2B can be open to all interested parties (e.g. commodity exchange) or limited to specific, pre-qualified participants (private electronic market). EC that is conducted between businesses and consumers, on the other hand, is referred to as business-to-consumer or B2C. This is the type of EC conducted by companies such as Amazon.com.

EC is generally considered to be the sales aspect of e-business. It also consists of the exchange of data to facilitate the financing and payment aspects of the business transactions.

This paper surveys EC technology trends and forecasts using bibliometric analysis from 1989 to 2009 with topic as "EC" in SSCI database. The bibliometric analytical technique is used to examine the topic in SSCI journals from 1989 to 2009, based on the scope of 2655 literatures of EC. This paper implemented and classified EC literatures using the seven categories as: publication year, citation, country/territory, institute name, document type, language and subject area for different distribution status in order to explore how EC technology trends and applications have developed in this period.

There are concurrent research domains using bibliometric methodology to analyze the trend and forecasts, like ubiquitous computing, home networking, social capital, supply chain management, data mining, customer relationship management, and knowledge management. [1]-[9]

For verifying the analysis result, the paper will perform by the following steps to verify the reliability of Lotka's Law:

(1) Collect data

(2) List author & literature distribution table

(3) Calculate n value (slope)

(4) Calculate c value

(5) Utilizing Kolmogorov-Smirnov test, (K-S test) to evaluate if matched Lotka's Law [1]

The paper provides a roadmap to guide future research and abstract the trend information so that EC researchers can save some time since core knowledge will be concentrated in core categories. This implies that the phenomenon "success breeds success" is more common in higher quality publications.

2. Material and methodology

All documents used in this study were accessed from the database of the Social Science Citation Index (SSCI), obtained by subscription from the ISI, Web of Science, Philadelphia, PA, USA. In this study, we discuss the papers published in the period from 1989 to 2009 because there was no data prior to that year. The Social Sciences Citation Index is a multidisciplinary index to the journal literature of the social sciences. It fully indexes over 1,950 journals across 50 social sciences disciplines. It also indexes individually selected, relevant items from over 3,300 of the world's leading scientific and technical journals.

Bibliometrics is a type of research method used in library and information science. It utilizes quantitative analysis and statistics to describe patterns of publication within a given field or body of literature. Researchers may use bibliometric methods of evaluation to determine the influence of a single writer, for example, or to describe the relationship between two or more writers or works. One common way of conducting bibliometric research is to use the Social Science Citation Index (SSCI), the Science Citation Index (SCI) or the Arts and Humanities Citation Index (A&HCI) to trace citations.

There are some research domains using bibliometric methodology to analyze the trend and forecasts, such as ubiquitous computing, home networking, social capital, e-commerce, supply chain management, data mining and knowledge management [2]-[8].

2.1. Laws of Bibliometrics

One of the main areas in bibliometric research concerns the application of bibliometric laws. The three most commonly used laws in bibliometrics are: Lotka's law of scientific productivity, Bradford's law of scatter, and Zipf's law of word occurrence.

2.1.1. Lotka's Law

Lotka's Law describes the frequency of publication by authors in a given field. It states that "the number (of authors) making n contributions is about $1/n^2$ of those making one; and the proportion of all contributors, that make a single contribution, is about 60 percent" [9]. This means that out of all the authors in a given field, 60 percent will have just one publication, and 15 percent will have two publications ($1/2^2$ times 0.60). 7 percent of authors will have three publications ($1/3^2$ times 0.60), and so on. According to Lotka's Law of scientific productivity, only 0.6 percent of the authors in a field will produce more than 10 articles. Lotka's Law, when applied to large bodies of literature over a fairly long period of time, can be accurate in general, but not statistically exact. It is often used to estimate the frequency with which authors will appear in an online catalog [9].

Lotka's law is generally useful for understanding the productivity patterns of authors in a bibliography [10]-[15]. In this article, Lotka's Law is selected to perform bibliometric analysis to check on literature record count versus accumulated authors between 1989 and 2009 to perform author productivity inspection for discovering historical review and collecting the results for research tendency forecast in the near future. For verifying the analysis result, the paper implements K-S test to evaluate if the result matched Lotka's Law.

2.1.2. Bradford's Law

Bradford's Law serves as a general guideline to librarians in determining the number of core journals in any given field. It states that journals in a single field can be divided into three parts, each containing the same number of articles: (1) a core of journals on the subject, relatively few in number, that produces approximately one-third of all the articles, (2) a second zone, containing the same number of articles as the first, but a greater number of journals, and (3) a third zone, containing the same number of articles as the second, but a still greater number of journals. The mathematical relationship of the number of journals in the core to the first zone is a constant n and to the second zone the relationship is n². Bradford expressed this relationship as 1: n: n². Bradford formulated his law after studying a bibliography of geophysics, covering 326 journals in the field. He discovered that 9 journals contained 429 articles, 59 contained 499 articles, and 258 contained 404 articles. So it took 9 journals to contribute one-third of the articles, 5 times 9, or 45, to produce the next third, and 5 times 5 times 9, or 225, to produce the last third. As may be seen, Bradford's Law is not statistically accurate, strictly speaking. But it is still commonly used as a general rule of thumb [9].

2.1.3. Zipf's Law

Zipf's Law is often used to predict the frequency of words within a text. The Law states that in a relatively lengthy text, if you "list the words occurring within that text in order of decreasing frequency, the rank of a word on that list multiplied by its frequency will equal a constant. The equation for this relationship is: r x f = k where r is the rank of the word, f is the frequency, and k is the constant [9]. Zipf illustrated his law with an analysis of James Joyce's Ulysses. "He showed that the tenth most frequent word occurred 2,653 times, the hundredth most frequent word occurred 265 times, the two hundredth word occurred 133 times, and so on. Zipf found, then that the rank of the word multiplied by the frequency of the word equals a constant that is approximately 26,500" [9]. Zipf's Law, again, is not statistically perfect, but it is very useful for indexers.

3. Research finding and discussion

This research is accessing the Social Science Citation Index (SSCI) on Web of Science created by ISI. The result is summarizing those 2655 paper indexes which topic is "e-commerce" from 1989 to 2009, shown as Figure 1. Obviously, the literature production of EC is rising since 1997 and citation is also increasing steadily and gradually by every year. It shows the research of EC is very popular in the highly exploration period, referred to Figure 1. The research of EC reached the highest record in 2009.
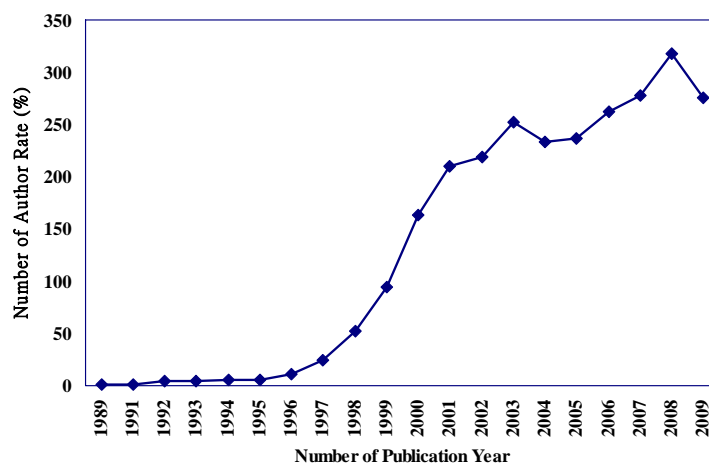

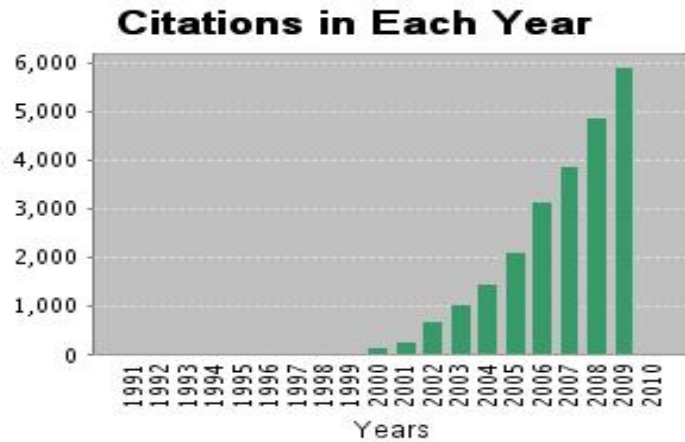**Figure 1.** The tendency chart of literature growth of EC

Figure 2. Citation of EC in each year (Source: SSCI database)

By viewing on Table 1, they displayed that the distribution of country/territory from 1989 to 2009, USA is a champion with 1346 record counts (50.70%), following by England, Taiwan, People R. China and Canada which achieved the record counts as 233(8.78%), 179(6.74%), 166(6.25%) and 135(5.08%) oppositely. For the distribution of institution name as shown in Table 2, the result indicates that USA is still the most productive country in the research aspect of EC in the world.

Table 1. Distribution of top 25 countries/territories from 1989 to 2009

Ranking	Country/Territory	NP	% of 2655
1	USA	1346	50.6968%
2	England	233	8.7759%
3	Taiwan	179	6.7420%
4	Peoples R China	166	6.2524%
5	Canada	135	5.0847%
6	South Korea	112	4.2185%
7	Australia	95	3.5782%
8	Germany	74	2.7872%
9	Netherlands	74	2.7872%
10	Spain	56	2.1092%
11	Singapore	54	2.0339%
12	Scotland	31	1.1676%
13	Italy	30	1.1299%
14	France	27	1.0169%
15	India	22	0.8286%
16	Finland	21	0.7910%
17	Greece	21	0.7910%
18	Switzerland	19	0.7156%
19	New Zealand	18	0.6780%
20	Wales	18	0.6780%
21	Israel	17	0.6403%
22	Ireland	15	0.5650%
23	South Africa	15	0.5650%
24	Sweden	14	0.5273%
25	Turkey	14	0.5273%

NP=number of publication

Table 2. Distribution of top 25 institutions from 1989 to 2009

Ranking	Institution name	NP	% of 2655
1	University of Minnesota	52	1.9586%
2	University of Maryland	45	1.6949%
3	University of Texas	43	1.6196%
4	University of Illinois	39	1.4689%
5	City University of Hong Kong	38	1.4313%
6	National University of Singapore	36	1.3559%
7	University of Pennsylvania	35	1.3183%
8	Georgia State University	34	1.2806%
9	Arizona State University	31	1.1676%
10	Michigan State University	31	1.1676%
11	Purdue University	27	1.0169%
12	University of Wisconsin	27	1.0169%
13	Drexel University	25	0.9416%
14	National Sun Yat Sen University	24	0.9040%
15	University of British Columbia	24	0.9040%
16	Carnegie Mellon University	23	0.8663%
17	University of Hong Kong	23	0.8663%
18	University of North Carolina	23	0.8663%
19	Chinese University of Hong Kong	22	0.8286%
20	Korea Advanced Institute of Science & Technology	22	0.8286%
21	Pennsylvania State University	22	0.8286%
22	University of California Irvine	22	0.8286%
23	Hong Kong Polytechnic University	21	0.7910%
24	University of Connecticut	21	0.7910%
25	Georgia Institute of Technology	19	0.7156%

NP=number of publication

Table 3. Distribution of document types from 1989 to 2009

Document type	NP	% of 2655
Article	2089	78.6817%
Proceeding Paper	206	7.7589%
Review	130	4.8964%
Editorial Material	116	4.3691%
Book Review	90	3.3898%
News Item	10	0.3766%
Letter	5	0.1883%
Meeting Abstract	5	0.1883%
Correction	2	0.0753%
Bibliography	1	0.0377%
Note	1	0.0377%

NP=number of publication

From the Table 3, it indicated that the most popular publication document type is article (2089 record counts, 78.68%), following by proceedings paper (206 record counts, 7.76%) and review (130 record counts, 4.8964%). The most popular language for literature is using English (2604 record counts, 98.08%) in the research domain of EC, following by German (26 record counts, 0.98%), see the following Table 4. The result shows that article document type and English language are still the main trend in EC research domain.

Table 4. Distribution of languages from 1989 to 2009

Language	NP	% of 2655
English	2604	98.0791%
German	26	0.9793%
Czech	5	0.1883%
French	4	0.1507%
Spanish	4	0.1507%
Rumanian	3	0.1130%
Slovak	3	0.1130%
Portuguese	2	0.0753%
Russian	2	0.0753%
Slovene	1	0.0377%
Swedish	1	0.0377%

NP=number of publication

Table 5. Distribution of subject areas from 1989 to 2009

Ranking	Subject area	NP	% of 2655
1	Management	764	28.78%
2	Computer Science & Information Systems	741	27.91%
3	Business	644	24.26%
4	Information Science & Library Science	576	21.69%
5	Operations Research & Management Science	301	11.34%
6	Computer Science & Interdisciplinary Applications	182	6.86%
7	Computer Science & Software Engineering	166	6.25%
8	Economics	163	6.14%
9	Engineering & Industrial,	137	5.16%
10	Computer Science & Artificial Intelligence	132	4.97%
11	Telecommunications	89	3.35%
12	Computer Science & Cybernetics	85	3.20%
13	Ergonomics	80	3.01%
14	Law	72	2.71%
15	Communication	62	2.34%
16	Business, Finance	55	2.07%
17	Geography	53	2.00%
18	Social Sciences, Interdisciplinary	51	1.92%
19	Psychology, Multidisciplinary	48	1.81%
20	Computer Science, Theory & Methods	44	1.66%
21	Psychology, Applied	43	1.62%
22	Planning & Development	41	1.54%
23	Environmental Studies	38	1.43%
24	Engineering, Electrical & Electronic	37	1.39%
25	Engineering, Manufacturing	36	1.36%

NP=number of publication

In the Table 5, it is important to summarize the trend information for EC researchers since core knowledge will be concentrated in core categories and to get understanding about the distribution of top 25 subject areas in future research trends and forecasts. Focus on the right categories, researchers will catch the core research information. The top three ranking of research domains are management (764 record counts, 28.78%), following by computer science & information systems (741 record counts, 27.91%) and business (644 record counts, 24.26%). Moreover, it also discovered that there are a lot of research domains for EC literature production such as information science & library science (576 record counts, 21.69%), operations research & management science (301 record counts, 11.34%), computer science & interdisciplinary applications (182 record counts, 6.86%), computer science, software engineering (166 record counts, 6.25%), economics (163 record counts, 6.14%), engineering & industrial (137 record counts, 5.16%) and computer science, artificial intelligence (132 record counts, 4.97%).

4. The literatures productivity analysis of EC by Lotka's Law

The section will perform by the following steps to verify the reliability of Lotka's Law:
(1)Collect data
(2)List author & literature distribution table
(3)Calculate n value (slope)
(4)Calculate c value
(5)Utilizing K-S test to evaluate if the result matched Lotka's Law

(1) Collect data and
(2) List author & literature distribution table:
It calculated the author quantity by the equality method from 2655 literatures which retrieved by index on SSCI. Thus, it is obtained altogether 1062 of authors on research aspect of EC. See the Table 6 for reference.

From Table 6, that would be a total of 1250 articles with 1062 authors with an average of 0.85 author for each article. The result indicates that the literatures of data mining were usually generated by single author.

Table 6. Distribution of author productivity of EC

NP	Author (s)	(NP) x (Author)	Accumulated Record	Accumulated Record %	Accumulated Author (s)	Accumulated Author (s)%
11	1	11	11	0.88%	1	0.09%
10	0	0	11	0.88%	1	0.09%
9	0	0	11	0.88%	1	0.09%
8	0	0	11	0.88%	1	0.09%
7	1	7	18	1.44%	2	0.19%
6	1	6	24	1.92%	3	0.28%
5	3	15	39	3.12%	6	0.56%
4	7	28	67	5.36%	13	1.22%
3	20	60	127	10.16%	33	3.11%
2	94	188	315	25.20%	127	11.96%
1	935	935	1250	100.00%	1062	100.00%

NP: number of publication

(3) Calculate n value (slope):
By the result of calculation on Table 7, it could bring into Lotka's Law's equation as below to calculate n value:

$$n = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2}$$

(1)

We can refer Table 7 for the values in equation,

$$n = \frac{11 * (2.06) - (7.6) * (7.57)}{11 * (6.3) - (7.6)^2}$$

(2)

Then n = -3.02918953

(4) Calculate c value:

After that, we also found c = 0.834721516, the equation is shown as below:

$$c = \frac{1}{\sum_1^{p-1} \frac{1}{x^n} + \frac{1}{(n-1)(p^{n-1})} + \frac{1}{2p^n} + \frac{n}{24(p-1)^{n-1}}}$$

(3)

P = 10 (max(x)-1),
x = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

When we got n = -3.02918953 and c = 0.834721516, it explored:

$$f(x) = 0.834721516/x^{3.02918953}$$

(4)

Table 7. Calculation of the exponent n for EC

x (NP)	y (Author)	X=log(x)	Y=log(y)	XY	XX
11	1	1.04	0.00	0.00	1.08
10	0	1.00	--	--	1.00
9	0	0.95	--	--	0.91
8	0	0.90	--	-	0.82
7	1	0.85	0.00	0.00	0.71
6	1	0.78	0.00	0.00	0.61
5	3	0.70	0.48	0.33	0.49
4	7	0.60	0.85	0.51	0.36
3	20	0.48	1.30	0.62	0.23
2	94	0.30	1.97	0.59	0.09
1	935	0.00	2.97	0.00	0.00
Total	1062	7.60	7.57	2.06	6.30

x= number of publication; y=author; X = logarithm of x; Y = logarithm of y

Referring to the datum from Table 6, authors with only one literature is 88.04% (100%-11.96%=88.04%), which is close to primitive c value 83.47% generated by Lotka's law. After that, it can follow the calculation to get n and c value by the least squares law, carry onto the further proceeding examination for Lotka's law compliance.

According to Pao [16] suggestion, the absolute value of n should be between 1.2 and 3.8 which formed by the generalized Lotka's Law, the result is matched the reference data by observation. The distribution chart is shown as Figure 3.

(5) Utilizing K-S (Kolmogorov-Smirnov, K-S) test to evaluate if matched Lotka's Law:

For discussing the value of n and c, we got n = -3.02918953 and c = 0.834721516 generated by Lotka's Law, the result demonstrated that the EC literature author distribution and the primitive Lotka's Law are matched approximately, refer to Figure 3. In order to examine whether the theoretical value and the observation value are tallied, the paper use K-S test to evaluate the suitability of Lotka's Law. Regarding the n and c value which gained by the formula, it is possible to calculate the expected value and the accumulation value of author, following by K-S test examination.
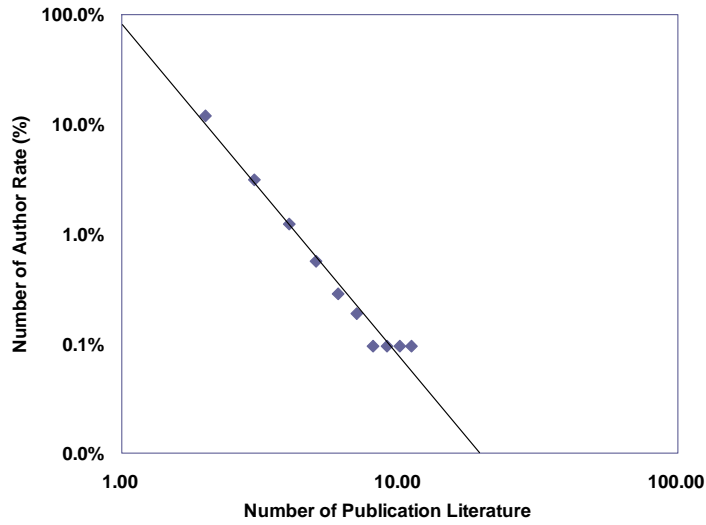
Figure 3. Distribution of literature productivity of author on EC research aspect

Table 8. The K-S test for EC

NP	Author (s)	EC (Observed)	Sn(x)	EC (Expected)	Fo(x)
1	0.8804	0.8804	0.8347	0.8347	0.0457
2	0.0885	0.9689	0.1023	0.9370	0.0320
3	0.0188	0.9878	0.0299	0.9669	0.0208
4	0.0066	0.9944	0.0125	0.9794	0.0149
5	0.0028	0.9972	0.0064	0.9858	0.0114
6	0.0009	0.9981	0.0037	0.9895	0.0086
7	0.0009	0.9991	0.0023	0.9918	0.0073
8	0.0000	0.9991	0.0015	0.9933	0.0057
9	0.0000	0.9991	0.0011	0.9944	0.0047
10	0.0000	0.9991	0.0008	0.9952	0.0039
11	0.0009	1.0000	0.0006	0.9957	0.0043

NP= number of publication; EC: author productivity of EC; Sn(x) = observed cumulative frequency; Fo(x) = theoretical cumulative frequency; D = maximum deviation

From Table 8, we can find D (D = Max | Fo(x) – Sn(x) |) = 0.0457. According to K-S test, the threshold value is:

$$1.63/\sqrt{1062} = 0.05 \tag{5}$$

Because Dmax is smaller than the K-S test threshold value, the result also indicated that the distribution of author productivity is matched by the Lotka's Law. The consequence means the Lotka's Law is suitable for the literature author productivity distribution in EC research domain [17].

5. Conclusion

EC is one of fast growing research topics in recently years, the trend forecast of this research field by each kind of literature characteristic and author productivity distribution are in growing period. In this analysis, it demonstrated that the current EC literatures are still continuously to grow, the main research development facility with delivered the largest production is USA, but England, Taiwan and Peoples R China, Canada and South Korea also have potential to deliver more literatures in the future. The paper is performed by the following steps to verify the reliability of Lotka's Law: (1) collect data; (2) list author & literature distribution table; (3) calculate n value (slope); (4) calculate c value; (5) utilizing K-S (Kolmogorov-Smirnov, K-S) test to evaluate if matched Lotka's Law. After checked by K-S test, the distribution of frequency indexes of author productivity is suitable for Lotka's Law. The distribution of author productivity consequence indicates that the literatures of EC were usually generated by single author. The most relevant disciplines for EC subject category provided by management, computer science and information systems, business, information science and library science, operations research and management science, computer science, interdisciplinary applications, computer science and software engineering, economics, engineering & industrial and computer science, artificial intelligence. Focus on the right categories, you will catch the core research information. The paper provides a roadmap to guide future research and abstract the trend information so that EC researchers can save some time since core knowledge will be concentrated in core categories. This implies that the phenomenon "success breeds success" is more common in higher quality publications.

6. References

[1] M. Y. Tsai, The characteristic of informetric and bibliometric, Hwa-Tai, Taiwan, 2003.

[2] C.Y. Yin & Y.P. Chi, "A Historical Review and Trend Analysis of Ubiquitous Computing Research Using Bibliometric Methodology During 1993 to 2009", International Journal of Digital Content Technology and its Applications (JDCTA), Vol. 4, No 2, 2010, pp.25~35.

[3] C.Y. Yin & J.M. Yang, "Assessment of Home Networking Research Trends, 1981-2009", Journal of Convergence Information Technology (JCIT), Vol. 5, No 2, 2010, pp.17~27.

[4] C.Y. Yin & J.K. Chiang, "Bibliometric Analysis of Social Capital Research During 1956 to 2008", Journal of Convergence Information Technology (JCIT), Vol. 5, No 2, 2010, pp.124~132.

[5] H.H. Tsai & Y.P. Chi, "Trend Analysis of Supply Chain Management by Bibliometric Methodology", International Journal of Digital Content Technology and its Applications (JDCTA), accepted and forthcoming in 2010.

[6] H.H. Tsai, "Analyzing Technology Trend and Forecasts of Data Mining: A Bibliometric Approach", International Journal of Digital Content Technology and its Applications (JDCTA), accepted and forthcoming in 2010.

[7] H.H. Tsai, & J.M. Yang, "Analysis of Knowledge Management Trend by Bibliometric Approach", In Proceeding(s) of the WASET on Knowledge Management, Vol. 62, pp. 1291~1296, 2010.

[8] H.H. Tsai, & J.M. Yang, "A Study of Data Mining Productivity Review from SSCI Database Using Bibliometric Methodology", In Proceeding(s) of the AIT on Advanced Information Technologies, 2010.

[9] W. G. Potter, " 'Of Making Many Books There is No End': Bibliometrics and Libraries", The Journal of Academic Librarianship, vol. 14, pp. 238a-238c, 1988.

[10] D. K. Gupta, "Lotka's law and productivity of entomological research in Nigeria for the period 1900–1973", Scientometrics, vol. 12, pp. 33–46, 1987.

[11] P. T. Nicholls, "Bibliometric modeling processes and empirical validity of Lotka's law", Journal of American Society for Information Science, vol. 40, no. 6, pp. 379–385, 1989.

[12] R. C. Coille, "Lotka's frequency distribution of scientific productivity", Journal of American Society for Information Science, vol. 28, pp. 366–370, 1977.

[13] J. Vlachy, "Frequency distribution of scientific performance: A bibliography of Lotka's law and related phenomena", Scientometrics, vol. 1, pp. 109–130, 1978.

[14] I. K. R. Rao, "The distribution of scientific productivity and social change", Journal of American Society for Information Science, vol. 31, pp. 111–122, 1980.

[15] M.L. Pao, "Lotka's law, a testing procedure", Information Processing and Management, vol. 21, pp. 305–320, 1985.

[16] M. L. Pao, Concept of Information Retrieve, Libraries Unlimited, Englewood, Colorado, 1989.

[17] W. G. Potter, "Lotka's Law Revisited", Library Trends, vol. 30, no. 1, pp.21-39, Summer 1981.

E-Commerce Research Trend Forecasting: A Study of Bibliometric Methodology
Hsu-Hao Tsai, Johannes K. Chiang
International Journal of Digital Content Technology and its Applications. Volume 5, Number 1, January 2011

- 111 -