

# Ontology: The Historical Review and Literature Productivity Analysis Using Bibliometric Methodology from 1956 to 2008

Chi-Yen Yin

National Chengchi University  
No.64, Sec. 2, Zhinan Rd., Wenshan  
District, Taipei City 11605, Taiwan  
+886-2-2264-8940

97356509@nccu.edu.tw

Yau-Jung Lee

National Chengchi University  
No.64, Sec. 2, Zhinan Rd., Wenshan  
District, Taipei City 11605, Taiwan  
+886-939-320-340

96356505@nccu.edu.tw

Jiann-Min Yang

National Chengchi University  
No.64, Sec. 2, Zhinan Rd., Wenshan  
District, Taipei City 11605, Taiwan  
+886-2-2939-3091

jmyang@mis.nccu.edu.tw

## ABSTRACT

The objective of this paper is mainly focusing on the historical review of international periodicals and literatures which topics as great renown of "Ontology" on SSCI database from 1956 to 2008. The result indicated that the literature productions related to ontology topic are still growing. The frequency indexes of author productivity distribution didn't follow by Lotka's Law. The applications of ontology are mainly following by research aspects such as philosophy; computer science and information system; information science and library science; psychology, multidisciplinary; history and philosophy of science; ethics and sociology and so on. The literatures of ontology are usually generating by multiple authorship.

## Keywords

Ontology, Literature productivity, Lotka's law

## 1. INTRODUCTION

Since PIOTROWSKI, ZA announced first paper at the journal of *NERVOUS AND MENTAL DISEASE* in 1956. The topic of ontology began a full exploration of all the possibilities in every research aspect in last fifteen years. This paper is applying bibliometric methodology toward onto literature productivity review and trend analysis which is also to get better understanding about the quantitative aspects of recorded information such as research authors, institutions, languages and subjects, finally proceeding by Lotka's law on papers versus authors between 1956 and 2008 to perform author productivity analysis, discovering historical tracking and collecting the results for research tendency forecasting in the near future.

"Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICIS 2009, November 24-26, 2009 Seoul, Korea  
Copyright © 2009 ACM 978-1-60558-710-3/09/11... \$10.00"

## 2. LITERATURE REVIEW

Ontology is the philosophical study of the nature of being, existence or reality in general, as well as of the basic categories of being and their relations. Traditionally listed as a part of the major branch of philosophy known as metaphysics, ontology deals with questions concerning what entities exist or can be said to exist, and how such entities can be grouped, related within a hierarchy, and subdivided according to similarities and differences. In computer science and information science, the ontology is a formal representation of a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, and may be used to define the domain. Ontology has recently gained importance in a variety of research aspects. In theory, the ontology is a "formal, explicit specification of a shared conceptualization". [1] The ontology provides a shared vocabulary, which can be used to model a domain — that is, the type of objects and/or concepts that exist, and its properties and relations. [2] Ontology is used in various research aspects such as artificial intelligence, the semantic web, software engineering, biomedical informatics, library science, and information architecture as a form of knowledge representation about the world or some part of it.

## 3. RESEARCH FINDING AND DISCUSSION

This research is utilizing the Social Sciences Citation Index (SSCI) of Web of Science created by ISI. The result is generating those 1,966 indexes of literature which titles are "Ontology" from 1956 to 2008 as well as parameters for next stage anticipation, shown as figure 1 and figure 2. Obviously, the literature production of ontology is rising up since 1995, and citation is also increasing steady and gradually by every year. It appears that the research of ontology is very popular and getting into the highly mature period in last fifteen years, it reached 1,000 citation times in 2005, and also over 2,000 citation times in 2008, please referred to figure 2. Moreover, there are a lot of research announcements from 2003 and 2008 which reached 5.14%, 4.88%, 6.87%, 8.9%, 8.7%, and 10.37% respectively.

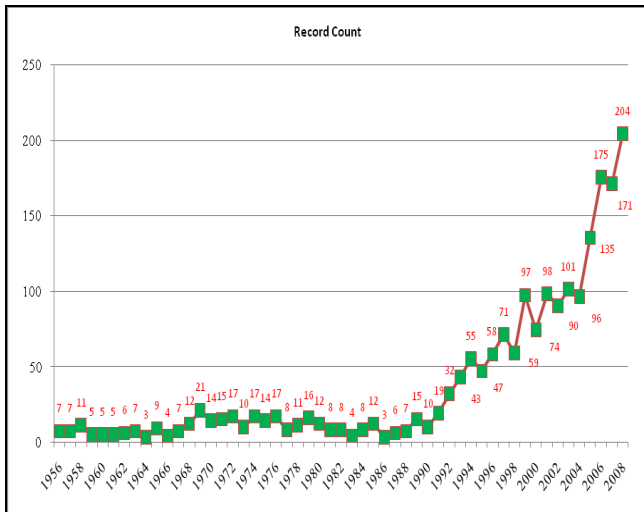


Figure 1, the tendency chart of literature growth of ontology

### Citations in Each Year

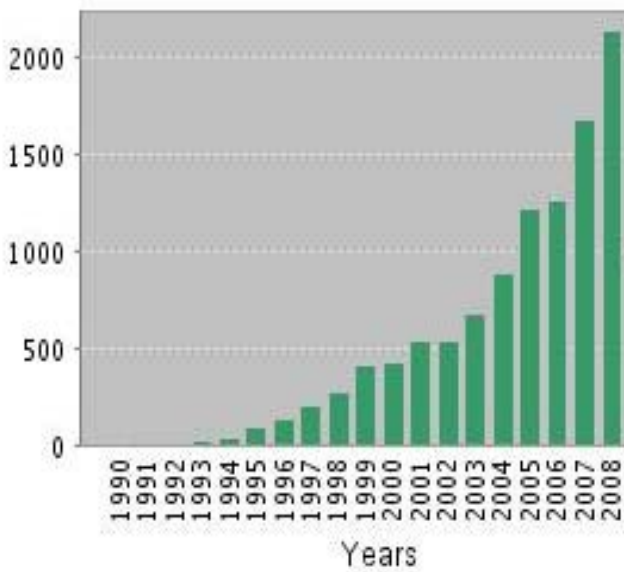


Figure 2, citation in each year (Source: SSCI database)

By viewing on figure 3, it displayed that the distribution of Country/Territory. United States is a champion, following by the England, Canada and Australia which achieved the record counts as 730(37.13%), 313(15.92%), 128(6.51%) and 86(4.37%) oppositely. Combining with the distribution of institutions name (as figure 4) for deeply observation, it shown that United States is still the most productivity country within the research aspects of ontology in the globe. However, the record counts are ranking as top research countries in Asia world such as Taiwan (22), Israel (18), Japan (14), China (13) and South Korea (8).

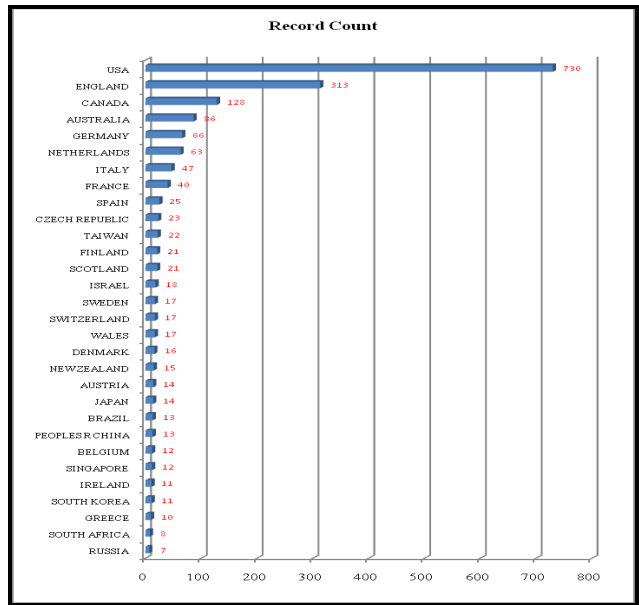


Figure 3, the distribution of country/territory from 1956 to 2008

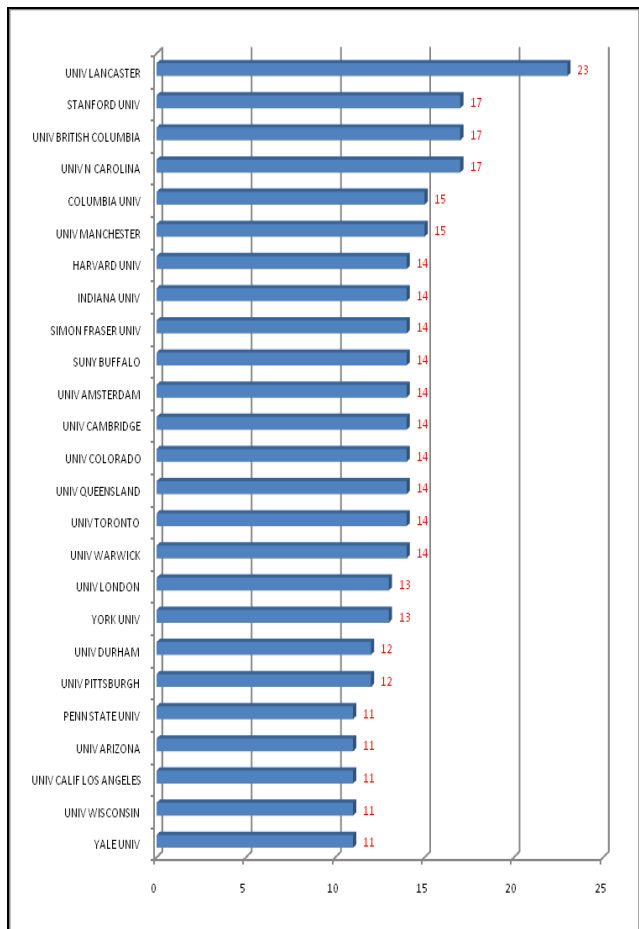


Figure 4, Distribution of Top 25 Institution Name from 1956 to 2008

On table 1, it indicated that the most publication document type is “Article” (1,282 record counts, 65.21%), and the most popular language for writing is using “English” (1,815 record counts, 92.32%) as well. See the following table 2.

**Table 1, Distribution of Document Type from 1956 to 2008**

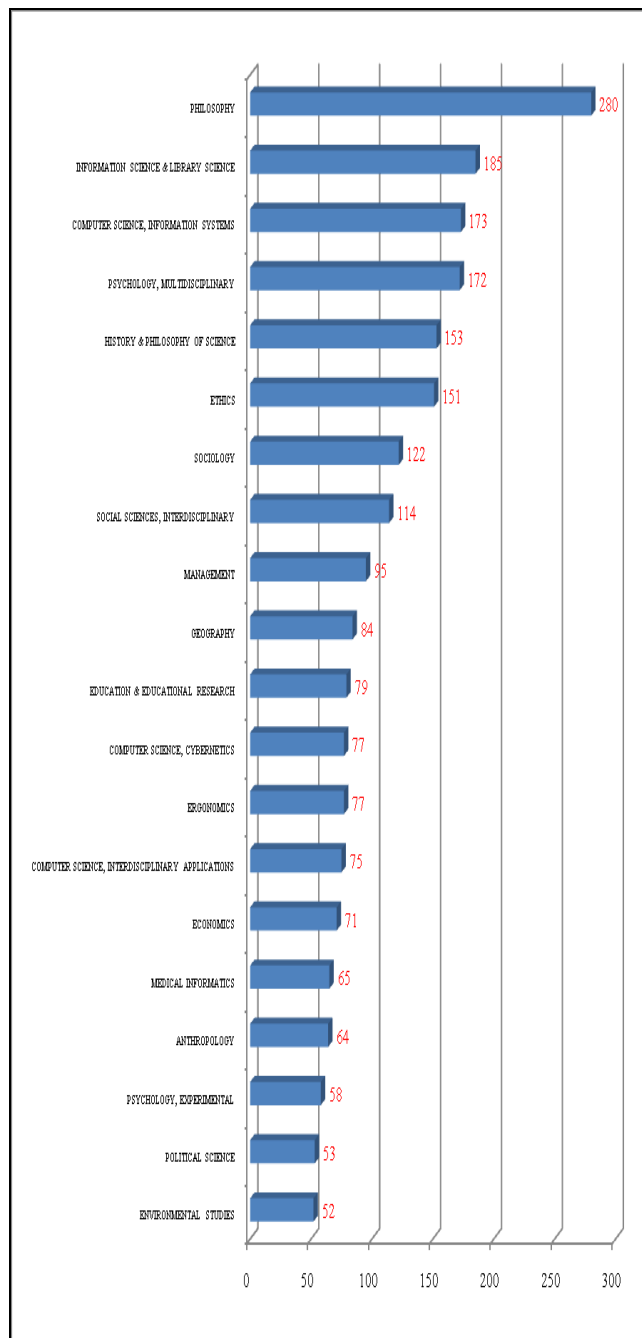
Document Type	Record Count	% of 1,966
ARTICLE	1,282	65.21%
BOOK REVIEW	257	13.07%
PROCEEDINGS PAPER	204	10.38%
REVIEW	89	4.53%
EDITORIAL MATERIAL	86	4.37%
MEETING ABSTRACT	14	0.71%
DISCUSSION	13	0.66%
NOTE	10	0.51%
LETTER	7	0.36%
CORRECTION	1	0.05%
CORRECTION, ADDITION	1	0.05%
ITEM ABOUT AN INDIVIDUAL	1	0.05%
REPRINT	1	0.05%

**Table 2, Distribution of Language from 1956 to 2008**

Language	Record Count	% of 1,966
ENGLISH	1,815	92.32%
GERMAN	48	2.44%
CZECH	47	2.39%
FRENCH	26	1.32%
SPANISH	11	0.56%
RUSSIAN	8	0.41%
DUTCH	6	0.31%
ITALIAN	1	0.05%
NORWEGIAN	1	0.05%
PORTUGUESE	1	0.05%
SLOVAK	1	0.05%
SWEDISH	1	0.05%

On the figure 5, it is to strengthen researchers to get understanding about the distribution of top 20 subject areas for future search and research directions. The top three ranking of research domains are philosophy (280 record counts, 14.24%), following by the information science and library science (185 record counts, 9.41%); computer science and information system (173 record counts, 8.80%). Furthermore, it also discovered that there are a lot of research aspects for ontology literature

production such as psychology and multidisciplinary; history and philosophy of science; ethics; sociology and social sciences and interdisciplinary and so on.



**Figure 5, the distribution of top 20 subject area from 1956 to 2008**

To sum up all of investigation on collected information, the research field of ontology is going to the exploded period. Most of literatures announced by United States, England, Canada and Australia, the United Nations positively appealed that governments and highly developed countries should invest massive resources within the research on the ontology which also carry onto the studies for each kind of phenomena such as nature,

society, and proposed the related discovery. It will be empowering the human knowledge in near future.

#### 4. The PRODUCTIVITY ANALYSIS OF LITERATURES OF ONTOLOGY BY LOTKA'S LAW

This section is mainly discussing the author to distribution situation of the rule which certificated by Lotka's law [2]. It is now calculating the author quantity by the equality method from 1,966 literatures which retrieved by index on SSCI. That is indicated that the degree of contribution of each author in one literature is the same, which could counts separately. Thus, it obtained altogether 2,690 of authors on Ontology research aspect. See the table 3.

**Table 3, the distribution of publication of author productivity of Ontology from 1956 to 2008**

Record Count	Author(s)	Paper Count	Accumulated Record Count	% of Accumulated Record Count	Accumulated Author(s)	% of Accumulated Author(s)
12	1	12	16	0.49%	1	0.04%
8	1	8	24	0.73%	2	0.07%
7	1	7	31	0.95%	3	0.11%
6	6	36	67	2.04%	9	0.33%
5	5	25	92	2.81%	14	0.52%
4	64	256	348	10.62%	78	2.90%
3	44	132	480	14.64%	122	4.54%
2	230	460	940	28.68%	352	13.09%
1	2338	2338	3278	100.00%	2690	100.00%

##### 4.1 Lotka's law

The research of discipline literature author distribution and productivity, may utilize the Lotka's law to discuss on it. The Lotka's law is called "a reverse square law of the scientific productivity", its connotation is: the number of author which published x literature is the number of author which published one literature total to divide x<sup>2</sup>. By performing Lotka's law to carry onto the analysis, which would be confirmed the literatures of Ontology whether to be suitable or not, it should also calculated the value of slope n, the value of constant c by using the examination of Kolmogorov-Smirnov test for determination whether the distribution is conform to or not (Prof. Dr. Tsay Ming-Yue, 2003). [3] Viewing on the datum, author has only 1 literature is 86.67%, which is not matched of primitive c value 60.79% provided by Lotka's law. After that, it can follow the calculation to get value of n and value of c by the least squares

law, carry onto the further proceeding examination for Lotka's law compliance.

Generally, Lotka's Law is an inverse square law that for every 100 authors contributing one article, 25 authors will contribute 2 articles, 11 authors will contribute 3 articles, and 6 authors will contribute 4 each. It may see a general decrease in performance among a body of authors following  $1:n^2$ . This ratio shows that some produce much more than the average which seems likely true for all kinds of content creation. However, Lotka doesn't take impact into account, only production numbers. Furthermore, in 1974, Voos found that in Information Science, the ratio was currently  $1:n^{3.5}$ . (Voos 1974) Thus, we can say that Lotka's Law may not be constant in value, but in following inverse square. Our challenge will then be to find the correct exponent in different mediums and fields. [6]

**Table 4, the productivity analysis of author and record count of Ontology from 1956 to 2008 – I**

Record Count	Author(s)	X=ln x	Y=ln y	XY	XX
12	1	2.48	0.00	0.00	6.17
8	1	2.08	0.00	0.00	4.32
7	1	1.95	0.00	0.00	3.79
6	6	1.79	1.79	3.21	3.21
5	5	1.61	1.61	2.59	2.59
4	64	1.39	4.16	5.77	1.92
3	44	1.10	3.78	4.16	1.21
2	230	0.69	5.44	3.77	0.48
1	2338	0.00	7.76	0.00	0.00
Total	2690	13.09	24.54	19.49	23.70

According to the result of calculation on table 4, it could bring into the following equation. The value of n = -3.477195

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

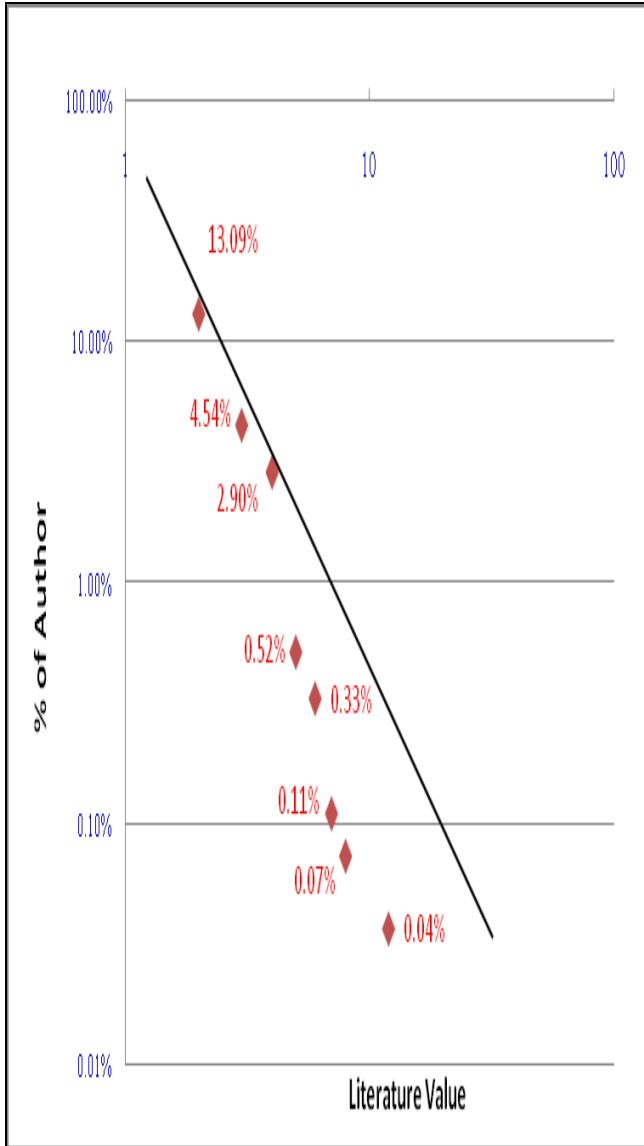
After that, we also found the value of c = 0.886663, the equation is shown as below:

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

p=8 (Max(x)-1), x=1, 2, 3, 4, 5, 6, 7, 8, 9,

when we got n = - 3.477195 , c = 0.886663, it explored, f(x)=c/x^n=0.886663/x<sup>3.477195</sup>. The distribution chart is shown as figure 6. While we discussing on the n and c value, primitive n approximately is -2, the value of c is 0.6079 which provided by Lotka's law, it demonstrated that the ontology literature author distribution and the primitive Lotka's law has not tallied completely. But actually it may observe the two datum

distribution disparity which not too big on figure 6. In order to examine the theoretical value and the observation value whether to tally. 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 Kolmogorov-Smirnov test examination. [4]



**Figure 6 : Distribution of literature productivity by author on Ontology research aspect**

According to Kolmogorov-Smirnov (K-S) test, table 5 demonstrated  $D_{max}=0.1052$ , but the sampling number is bigger than 35, therefore the value of threshold is  $1.63/2620^{1/2} = 0.031844685$ . Because  $D_{max}$  is much bigger than the threshold value, the result is this research author productive forces distribution and the Lotka's law does not tally, which means the Lotka's law is not suitable for ontology literature author productivity distribution datum.

**Table 5, the productivity analysis of author and record count of Ontology from 1956 to 2008 – II**

Record Count	Observation by Author	Accumulated Value	Expected Value by Author	Accumulated Value	ABS Value
		$S_n(X)$		$F_0(X)$	
1	0.8667	0.8667	0.886663973	0.791741	0.074959
2	0.0889	0.95559	0.079619427	0.87136	0.084228462
3	0.0280	0.98361	0.019440856	0.89080	0.092806929
4	0.0068	0.99037	0.007149555	0.89795	0.092420659
5	0.0048	0.99520	0.003290819	0.90124	0.093960758
6	0.0029	0.99810	0.001745723	0.90299	0.095113585
7	0.0010	0.99907	0.001021382	0.90401	0.095058386
8	0.0010	1.00003	0.000642006	0.90465	0.095382564
12	0.0010	1.00100	0.00015676	0.90481	0.096191988

## 5. CONCLUSION

Ontology is one of most popular discussion subjects in recent years, this historical review and trend forecast of this research field by each kind of literature characteristic and author productivity distribution is getting in highly mature period, it might be knowing that the present ontology literatures is still continually to grow, the main research development facility with delivered the large production is United States, but England, Canada, Australia, even some Asia countries such as Taiwan, Japan and South Korea, these non-US individual authors literature delivered actually are also very popular. The frequency indexes of author productivity distribution didn't follow by Lotka's Law. The applications of ontology are mainly following by research aspects which in term of philosophy; computer science and information system; information science and library science. The literatures of ontology are usually generating by multiple authorship.

## 6. REFERENCES

- [1] Tom Gruber (1993). "A translation approach to portable ontology specifications", In: *Knowledge Acquisition*. 5: 199-199.
- [2] Fredrik Arvidsson and Annika Flycht-Eriksson, *Ontologies I*, Retrieved 26, Nov 2008.
- [3] Prof. Dr. Tsay Ming-Yue (2003), *The Characteristic of Informetrics and Bibliometrics*, Hwa-Tai bookstore, Taiwan.
- [4] Prof. Dr. Tsay, Ming-Yue (1987), *The literature bibliometrics*, *Journal of Education and Library science*, Taiwan, Vol. 24, 3, pp. 261-296.
- [5] Prof. Dr. Tsay, Ming-Yue (Dec., 1998), *Bradford's Law*, *Journal of Information Communication and Library science*, Taiwan, Vol. 5, 2, pp. 25-49.
- [6] Turnbull, Don, 2005, 4, "Bibliometrics and the World-Wide Web", <http://www.ischool.utexas.edu/~donturn/research/bibweb.html>