2011 International Conference on Information Communication and Management
IPCSIT vol.16 (2011) © (2011) IACSIT Press, Singapore

Information Use by Scholars in Bioinformatics: A Bibliometric View

M.Nagarajan[1], L.Mohamed Idhris[2], P.Chellappandi[3], J.P.S.Kumaravel[4] and V.Premalatha[5]

[1]Professor & Head, Department of Library & Information Science, Annamalai University, Chidambaram – 608002. E-mail: au_nagarajan@yahoo.co.in

[2]Research Scholar, DOLIS, Madurai Kamaraj University, Madurai – 625 021. E-mail: lmidhris@gmail.com

[3]Assistant Professor, Department of Library & Information Science, Madurai Kamaraj University, Madurai – 625 021. E-mail: pandi_chella@rediffmail.com

[4]Deputy Librarian, Dr.T.P.M.Library, Madurai Kamaraj University, Madurai – 625 021. E-mail: jpskumar@yahoo.com

[5]Librarian, Thiagarajar School of Preceptors, Teppakulam, Madurai – 625009. E-mail: ptm.premalatha@gmail.com

Abstract. This paper reports on an analysis of the use of literature by scholars in the field of bioinformatics with the objective to identify the age of the information used and the type of journals referred for the research. Our results present the main journals used for bioinformatics research.

Keywords: Bibliometric, Bioinformatics.

1. Introduction

Bioinformatics is the application of computer technology to the management of biological information. Computers are used to gather, store, analyze and integrate biological and genetic information which can then be applied to gene-based drug discovery and development. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned. Important sub-disciplines within bioinformatics and computational biology include:

- the development and implementation of tools that enable efficient access to, and use and management of, various types of information
- the development of new algorithms (mathematical formulas) and statistics with which to assess relationships among members of large data sets, such as methods to locate a gene within a sequence, predict protein structure and/or function, and cluster protein sequences into families of related sequences

In all disciplines, knowledge is built by responding to the ideas and discoveries of those who came before us, Scholarly or academic journals are essential resources for doing academic research. Scholarly journal articles are unique in that they require authors to document and make verifiable the sources of the facts, ideas, and methods they used to arrive at their insights and conclusions. Scholarly articles also strive to identify and discuss the merits of alternative explanations and viewpoints for the positions they espouse. This makes it easier to assess the truth, as well as the strengths and weaknesses, of the claims made in a paper.

Research feeds on other research. A researcher will typically review the available literature on a topic of interest prior to launching into a further investigation or into a new investigation. This process of examining existing literature in an area of interest is know as a literature review. The process of reviewing the literature helps the researcher understand what else needs to be discovered about the chosen topic and uncovers gaps in knowledge about the topic that the current research can then attempt to fill in. In some cases, a writer may do just the literature review to inform other experts in the field of the current state of knowledge. In most cases,

however, the literature review is the launching pad from which a researcher proceeds into an original study of other issues or aspects of the topic.

This paper reports on the referencing practices of bioinformatics researchers using bibliometric techniques. A search was designed to download the research productivity in Bioinformatics from Thomson's ISI Web of Science Database. The downloaded data of 8729 articles have 313341 references.

2. Bibliometrics

Since the early 1900s, bibliometrics have been used to measure the impact of publications, scientists and journals (Godin, 2006). Bibliometrics are the measurement of text and information in documents and document related processes (Borgman and Furner, 2002). Originally, psychologists started counting publications systematically. Since the 1950s bibliometrics have become a broader scholarly activity to measure scientific output (Garfield, 1955).

3. Previous Studies

Lisee et al. (2008) collected all references from papers in the WoS indexed journals, from 1980 to 2005, and counted how many of those references were to papers published in proceedings. Overall, only 1.7% of the Natural Sciences and Engineering references were to proceedings, and 2.5% for the Social Sciences and Humanities, and these proportions were decreasing in time. But within Natural Sciences and Engineering, Computer Science papers had on average 19.8% of their references to proceedings. CS is the subfield with the highest of such proportion, followed by Electrical Engineering and Electronics (13.1%).

Lisee et al. (2008) also analyzed the age of the destinations, that is, the time interval between the year the source was published and year the document cited in the reference was published. For the whole set of destinations, the median age (or half-life) is 4.0 years for proceedings references, and 6.1 for all references. For Engineering (which includes CS), the average age of all references was 10.9 years, but the average age cited proceedings papers was only 7.1 years. Thus, the age data corroborates the idea that conference papers are more representative of forefront scientific results, which will obsolesce faster.

4. Methods

The methodology used was similar to other bibliometric studies published from our research team (Bliziotis et al., 2005; Falagas et al., 2005, 2006; Rosmarakis et al., 2005;Vergidis et al., 2005). The number of published articles was considered as an index of quantity of research productivity. The uniqueness of this research is the application of bibliometric tools for the references appended to research articles in Bioinformatics. For the purposes of our study we analyzed data from Thomsons ISI Web of Science Database for the period from 2000 to 2010. We computed some descriptive statistics such as the distribution of number of references per paper and the distribution of the ages of the cited documents in the source data set.

5. Analysis and Results

The research productivity in the field of Bioiformatics shows a growing trend from 2001 to 2010. Though there is growth in research, the growth rate varies over the years. The growth rate is maximum in the year 2002 and minimum in the year 2008. Hence it can be found that the growth of research in bioinformatics is not uniform over the period of eleven years.

Table 1 Frequency distribution of research productivity in Bioinformatics

Year	Count	Growh Rate
2000	163	
2001	214	0.31
2002	376	0.76
2003	556	0.48
2004	737	0.33
2005	927	0.26
2006	993	0.07
2007	1056	0.06
2008	1078	0.02
2009	1213	0.13
2010	1416	0.17
	8729	

Biglu (2007) reports that the average number of references per document for all papers indexed by Thomson Reuters Science Citation Index (SCI) for 2005 was 34.63. Vieira and Gomes (2010) also uses SCI data for 2004, and reports on average 18.25 and 24.11 reference per articlefor Mathematics and Physics respectively. In the present study the mean number of references per publication works out to 33.95 as shown in table 2.

Table 2 Number of references per article

Year	Publications	References	Ref/Pub
2000	163	4593	28.18
2001	214	6025	28.15
2002	376	11851	31.52
2003	556	17658	31.76
2004	737	24100	32.7
2005	927	30314	32.7
2006	993	34139	34.38
2007	1056	38133	36.11
2008	1078	40979	38.01
2009	1213	46484	38.32
2010	1416	58944	41.63

The mean age of references to articles also varies in various years. The total number of references appended to 8729 articles is 313341 of which only a negligible references are more than 100 years old. Further it is to be noted that nearly one third of the references are more than 10 years old and this is shown in table 3

Table 3 Age of References

Year	References	Percent
Before 1900	89	0.03
1900-1999	106169	33.88
2000-2010	207083	66.09
	313341	100

An analysis of the age of the references to articles in the subject Bioinformatics published in the years between 2000 and 2010 shows that references of age 10 and less than 10 years is more in number in all the years. As the age of the references increases over 10, the number of references decreases. This shows that scholars in Bioinformatics refer journal articles of recent years within 10 years as revealed in Table 4.

Table 4 Trend of Age of References

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
10 Year Ref	3818	5142	9969	14617	19908	24504	26954	29530	31274	35188	43443
20 Year Ref	511	640	1394	2262	3141	4272	5383	6563	7277	8504	11292
30 Year Ref	170	161	320	488	687	948	1132	1287	1496	1736	2512
40 Year Ref	57	55	99	173	232	387	428	460	580	638	853
50 Year Ref	21	13	31	51	63	97	138	144	207	212	462
above 50 Year Ref	16	14	38	67	69	106	104	149	145	206	382

The highly referred journals are given in Table 5. The most highly referred journal is Nucleic Acid Research whose frequency is 17009 and the second ranked referred journal is Proceedings of National Academy of Sciences, USA. Bioinformatics is in the third place. All the highly referred journals by Bioinformatics scholars have high impact factor.

Table 5 Highly Referred Journals

Sl.NO	Journals	Count
1	NUCLEIC ACIDS RES	17009
2	P NATL ACAD SCI USA	11483
3	BIOINFORMATICS	9492
4	NATURE	8518
5	SCIENCE	8175
6	J BIOL CHEM	8123
7	J MOL BIOL	7535
8	GENOME RES	3781
9	CELL	3767
10	NAT GENET	3620
11	PROTEINS	3025
12	INT J ONCOL	3019
13	BMC BIOINFORMATICS	2827
14	BIOCHEM BIOPH RES CO	2655
15	J BACTERIOL	2448
16	INT J MOL MED	2218
17	CANCER RES	2192
18	BIOCHEMISTRY-US	1965
19	J VIROL	1958
20	EMBO J	1920
21	NAT BIOTECHNOL	1904
22	GENOME BIOL	1774
23	PROTEIN SCI	1736
24	PROTEOMICS	1731
25	FEBS LETT	1679

Table 6 shows the most prolific authors referred by Bioinformatics scholars. The name of the most prolific author is Katoh, M from Japan. Here it is interesting to note that Katoh, M has published 134 papers in Bioinformatics during the period of 11 years from 2000 to 2010 and he is the most referred author by other scholars during this period. Hence Katoh, M canbe called as Author Classic in the field of Bioinformatics.

Table 6 Most Prolific Authors referred

Sl.NO	Author Name	Total Count
1	Katoh M	4127
2	ALTSCHUL SF	1618
3	Chou KC	1022
4	Kirikoshi H	716
5	Saitoh T	711
6	THOMPSON JD	678
7	Ashburner M	460
8	Bairoch A	432
9	Berman HM	387
10	Kanehisa M	352
11	ROST B	347
12	Katoh Y	312
13	Jones DT	310
14	BATEMAN A	310
15	PEARSON WR	302
16	EISEN MB	282
17	FELSENSTEIN J	256
18	Lander ES	247
19	Benson DA	240
20	Krogh A	239
21	EDDY SR	235
22	SAMBROOK J	216
23	HOLM L	216
24	Wheeler DL	205
25	SMITH TF	197

6. Conclusion

Referencing is a system used in the academic community to indicate where ideas, theories, quotes, facts and any other evidence and information used to support your assignments, can be found. Because of the level of authority and credibility evident in scholarly sources, they contribute a great deal to the overall quality of your papers. Use of scholarly sources is an expected attribute of academic course work. As discussed in this paper, the average number of references to a research paper in Bioinformatics is 34 and the

ideal age of references is below 10 years. This seems to indicate that some Bioinformatics research is either inspired by or try to argue its relevance based on current research

7. References

[1] Godin, B., 2006. *On the origins of bibliometrics1*. Scientometrics 68, 109–133.

[2] Borgman, C.L., Furner, J., 2002. *Scholarly communication and bibliometrics*. Annual Review of Information Science and Technology 36, 3–72.

[3] Garfield, E., 1955. *Citation indexes for science: a new dimension in Documentation through association of ideas*. Science 122, 108–111.

[4] Bliziotis, I.A., Paraschakis, K.,Vergidis, P.I., Karavasiou, A.I., Falagas, M.E., 2005.*Worldwide trends in quantity and quality of published articles in the field of infectious diseases*. BMC Infect. Dis. 5, 16.

[5] Falagas, M.E., Karavasiou, A.I., Bliziotis, I.A., 2005. *Estimates of global research productivity in virology*. J. Med. Virol. 76, 229–233.

[6] Biglu, M. H. (2007). *The influence of references per paper in the sci to impact factors and the Matthew effect*. Scientometrics, 74(3), 453–470.

[7] Vieira, E., & Gomes, J. (2010). *Citations to scientific articles: Its distribution and dependence on the article features*. Journal of Informetrics, 4, 1–13.

[8] Lisee, C., Lariviere, V., & Archambault, E. (2008). *Conference proceedings as a source of scientific information: A bibliometric analysis*. Journal of the American Society for Information Science and Technology, 59(11).