Scientometrics
DOI 10.1007/s11192-011-0412-z

Article title type and its relation with the number of downloads and citations

Hamid R. Jamali · Mahsa Nikzad

Received: 19 April 2011

Abstract Title of an article can be descriptive, declarative or a question. It plays important role in both marketing and findability of article. We investigate the impact of the type of article titles on the number of citations and downloads articles receive. Number of downloads and citations for all articles published in six of PLoS (Public Library of Science) journals (2,172 articles) were obtained from PLoS and type of each article's title (including descriptive, indicative and question) was determined as well as the number of substantive words in title (title length). Statistical difference and correlation tests were carried out. The findings showed that differences exist between articles with different types of titles in terms of downloads and citations, especially articles with question titles tended to be downloaded more but cited less than the others. Articles with longer titles were downloaded slightly less than the articles with shorter titles. Titles with colon tended to be longer and receive fewer downloads and citations. As expected, number of downloads and citations were positively correlated.

Keywords Article titles · Download · Citation · Title length · Colon

Introduction

Title is the most important element of any scientific article and the main indication of article's subject. Title is used by users as the main source of information for doing relevance judgment at the time of literature searching. Titles are very helpful for those who scan in libraries, catalogues, periodical indexes, references, databases, and tables of

H. R. Jamali (✉)
Department of Library and Information Studies, Faculty of Psychology and Education, Tarbiat Moallem University, No 49, Mofateh Ave, P.O. Box 15614, Tehran, Iran
e-mail: h.jamali@gmail.com

M. Nikzad
Iranian Research Institute for Information Science and Technology (IRANDOC), Information Science Research Center, No. 1090, Enghelab Ave., Felestin Cross, 1315773314 Tehran, Iran
e-mail: nikzad.mahsa@gmail.com

Published online: 19 May 2011

contents of edited books, reports, and proceedings (Soler 2007). Titles also play an important role in the marketing of the article as hundreds of articles are published every year in every scientific discipline and they in a sense compete with each other to be read.

The title of an article precedes the beginning and should clearly indicate the subject and arouse interest. An ideal title should be rather short, informative and attractive. But as Kane (2000) suggested it is difficult to balance these qualities, and most titles appear to be attractive but not informative, or informative but not attractive. Each title should motivate the reader to read an article; give the readers a summary of the contents, an overview of the topics and findings discussed; introduce the way in which the reported items are looked at (Ball 2009) and draw readers' attention to a article by informing them of its contents (Manten and Greenhalgh 1977). However, many authors fail to choose a suitable title for their articles (Manten and Greenhalgh 1977).

Therefore, as Wang and Bai (2007) maintained, the quality of titles can affect research articles' impact. An informative and attractive title might lead to more use, which in turn is manifested in the number of downloads or citations the article receives. There are quite a number of factors that could affect the impact of an article, including the significance and availability of the journal in which it is published, publication type, its subject, its author(s), its length and so on. Title is also one of these factors and there might be an association between title contents and syntax and subsequent article citations and downloads.

We aim to find out whether there is any link between the type of article title and its use (download and citation) by studying articles published in six of PLoS journals in 2007. We further aim to find out about any impact that the presence of colon, number of title words and article length can have on downloads and citations.

Types of title

We can consider three main types of titles.

- Declarative title: These titles include what articles say (their main conclusions), not just what they cover, for instance: "Adipose gene expression prior to weight loss can differentiate and weakly predict dietary responders".
- Descriptive or neutral titles: A descriptive title only describes the subject of the articles and does not reveal the main outcome or conclusion. For example: "A worldwide phylogeography for the human X chromosome".
- Interrogative (question) titles: These titles indicate the subject of the article in the form of a question that appeal to the curiosity of readers (Gustavii 2008). "Are there rearrangement hotspots in the human genome?"

Besides the above three main types of title, we can also find compound titles. For example, a title may start with a short question followed a subject sentence, or it may start with a noun phrase followed by a colon and a declarative sentence or a question. However the aforementioned three types are the main ones.

From the perspective of academic writing and scholarly communication, the most recommended type of title is declarative title which conveys the largest amount of information to user. However, not in every single article it is possible to form a declarative sentence as the title and in some fields (such as medicine) editors are more cautious in accepting declarative titles because of the impact they could have on the public health. Use of question mark in titles are not recommended generally (Gustavii 2008), however, their

use has been in steady rise during the last four decades (Ball 2009), in some fields more than the others. Interrogative titles are more acceptable for review articles than research articles (e.g., How long is a giant sperm? (Gustavii 2008). The most common type of title still seems to be descriptive title.

Literature review

Article titles have been studied from different perspectives including linguistic, scientometric and scholarly communication. Yitzhaki (1997) compared informativity (number of substantive words) of titles in humanities articles with sciences and social sciences between 1940 and 1960 and found significant differences. Yitzhaki (1994, 2002) looked at the length of articles and found a positive correlation between the title length, number of authors and the length of the article. Lewison and Hartley (2005) examined the number of words and the presence of colon in article titles over time and found that they both have increased from 1981 to 2001. Wang and Bai (2007) studied the syntax of 417 article published in *New England Journal of Medicine* (2003–2005) and found that nominal groups were widely used (in 99% of titles). Soler (2007) compared the syntax of titles in English articles with Spanish articles. Some other studies such as Anthony (2001) and Wang and Bai (2007) studied title syntax and issues such as term frequencies and their relation with citation rates. A few studies also examined the predictive power of titles for evaluating future citations (Sagi and Yechiam 2008; Jacques and Sebire 2009).

Sagi and Yechiam (2008) examined the association of humor in scientific article titles with the number of citations of an article for two psychology journals. They also examined the association between the levels of amusement and pleasantness and the article's monthly citation average. Their study showed that while the pleasantness rating was weakly associated with the number of citations, articles with highly amusing titles received fewer citations. The negative association between amusing titles and subsequent citations could not be attributed to differences in the title length and pleasantness, number of authors, year of publication, and article type.

In a large-scale study, Ball (2009) examined the titles of nearly 20 million scientific articles in Scopus in physics, life sciences and medicine from 1966 to 2005. He found a significant increase from 50% to more than 200% in the number of articles with question-mark titles over the 40 years. He mentioned 'marketing aspect' as one of the decisive factors behind the growing usage of question-mark titles in scientific articles. The presence of colon in titles and its relation with citation rate has been investigated by Hartley (2007a, 2007b). He showed that there were variations in different disciplines in terms of the use of colon in article titles and maintained that the use of colon has no effect upon their subsequent citation rate.

The most relevant study to ours is the one by Jacques and Sebire (2009). They reviewed the title characteristics of the 25 most cited articles and the 25 least cited articles published in 2005 in general and specialist medical journals (the *Lancet*, *BMJ* and *Journal of Clinical Pathology*). They found that the length of the title, the presence of a colon and an acronym in the title is positively correlated with the number of citations. Factors that caused poor citation included reference to a specific country in the title. Some of their findings were later questioned by Moore (2010), editor-in-chief of a biomedical journal. He, in a blog post, argued against Jacques and Sebire and used some data to show that their findings might be field-specific. He also argued that the correlation of citations and title length in some cases are negative.

The review of the literature shows that although the content and the syntax of article titles have been studied from different perspectives, no study has yet focused on the possible relation between the type of title and its downloads or citations. Moreover, no study yet investigated the relation between titles' characteristics and downloads. We aim to explore this less studied area.

Method

We downloaded the list of all articles published in 2007 in six journals of the Public Library of Science (PLoS) from Scopus that included the number of citations to each article. PLoS publishes seven open access electronic journals in the fields of medical and life sciences. We used the data for six of them (as follows) and ignored *PLoS Neglected Tropical Diseases* as its publication started in October 2007 and did not have a full volume in that year.

- PLoS One, 1227
- PLoS Biology, 223
- PLoS Medicine, 137
- PLoS Computational Biology, 202
- PLoS Genetics, 211
- PLoS Pathogens, 175

We chose 2007 for we thought that at least three year would be needed for articles to receive citations. We chose PLoS journals because they are prestigious journals in their own fields, they are open access and most of all, PLoS platform is one of the few journal platforms that give accurate up-to-date download data at article level. We limited our study to 'Articles' (i.e., research papers) and removed all of the editorials, errata, letters, reviews and short surveys from the data. In total, 2172 articles were included in the study.

The data were downloaded as comma delimited files from Scopus in November 2010 and imported into an Excel file. We read all of the titles one by one in a time-consuming process and for each title we determined the number of substantive words, presence of colon, and its type. For counting words, we used a stop list to the non-informative words (articles, prepositions, conjunctions, pronouns and auxiliary verbs), following Tocatlian (1970); Buxton and Meadows (1977), and Diodato (1982). We also counted abbreviations as words.

We then obtained download data for each article using the data available on PLoS website. We obtained the data on March 2011 and data included download statistics from 2007 up to the end of October 2010. PLoS gives four download statistics including the total number of downloads and the number of downloads for PDF, HTML, and XML formats separately. We used total download for our analysis.

Non-parametric tests were used as the data were not normally distributed.

Results

Title type, downloads and citations

A Kruskal–Wallis test was conducted to evaluate differences among the three title types (indicative, descriptive and question) on median change in number of downloads. The test

was significant $\chi^2(df = 2;\ N = 2{,}147) = 18.04,\ p < .001$. The summary presented in Table 1 may be needed to better understand these results. Follow-up Mann–Whitney U tests were conducted to evaluate pairwise differences among the three types of titles, controlling for Type I error across tests by using the Bonferroni correction. The results of these tests indicated a significant difference between the numbers of downloads for articles with question-type titles ($Mdn = 3{,}723$) and articles with declarative titles ($Mdn = 2{,}565$; $U = 10292.5$, $P < 0.001$; $r = 0.14$). Articles with question-type titles were downloaded more than articles with any of the other two types of titles. Articles with descriptive titles were also downloaded more than articles with declarative titles. Details (values of U, P, and r as a measure of effect size, $r = \frac{Z}{\sqrt{N}}$) of pairwise Mann–Whitney tests are presented in Table 2. It should be said that Bonferroni adjustment is said to be too conservative (Wu and Chen 2000) and while all P values are smaller than 0.005 they cannot be considered as significant values because Bonferroni adjusted significance in all pairwise comparisons should be smaller than the significant result on the overall test i.e., they should be at least 0.001.

Similar tests and procedure were carried out for citation data and the results of Kruskal–Wallis test was significant χ^2 ($df = 2$; $N = 2{,}147$) $= 10.66$, $p = 0.005$. Follow-up Mann–Whitney U tests for the evaluation of pairwise differences were also carried out using the Bonferroni method to control Type I error. The results showed significant differences between the numbers of citations received by articles with question-type titles ($Mdn = 6$) and those with descriptive titles ($Mdn = 14.23$; $U = 23{,}822$, $P = 0.002$; $r = 0.079$) and declarative titles ($Mdn = 12$; $U = 10613.5$, $P = 0.001$; $r = 0.12$). Table 3 relates.

Title length, title type, downloads and citations

Table 4 shows summary statistics for the number of words for different types of titles and we can see that declarative titles tend to be longer than the other two title types. This is not surprising as declarative titles are full sentences with verbs and writing a full sentence normally needs more words compared to noun phrases. Paired Mann–Whitney U tests showed that the difference between the number of words of declarative titles and the other two titles was statistically significant at $P < 0.001$.

Spearman correlation tests (Table 5) showed a very weak and negative correlation between number of downloads and title length meaning that articles with longer titles received relatively fewer downloads. However, no significant correlation was found between title length and citations. Numbers of downloads and citations were positively correlated ($rho = 0.55$; $P < 0.001$). Articles with longer titles seemed to be slightly longer themselves ($rho = 0.123$; $P < 0.001$).

Table 1 Number of downloads and citations for articles with different types of title

Title type	No	Download		Citation	
		Mean	Median	Mean	Median
Descriptive	1,442	3,906	2,754	16.92	14.23
Declarative	660	3,588	2,565	16.93	12
Question	45	5,817	3,723	10.47	6

Table 2 Mann-Whitney *U* test results for pairs of title types and downloads

		Declarative	Question
Descriptive	*U*	446697.5	25394
	P	0.004	0.004
	r	0.062	0.073
Declarative	*U*	–	10292.5
	P	–	0.000
	r	–	0.14

Table 3 Mann-Whitney *U* test results for pairs of title types and citations

		Declarative	Question
Descriptive	*U*	4,63,668	23,822
	P	0.345	0.002
	r	0.02	0.079
Declarative	*U*	–	10613.5
	P	–	0.001
	r	–	0.12

Table 4 Summary statistics for number of words in different types of titles

	Mean	Median	*SD*
Descriptive	8.73	8	2.69
Declarative	10.22	10	2.65
Question	8.54	8	3.06

Table 5 Correlation coefficients for pairs of variables

		Title length	Download	Times cited
Download	Spearman *rho*	−0.183		
	Sig.	0.000		
	N	2,169		
Times cited	Spearman *rho*	−0.022	0.550	
	Sig.	0.304	0.000	
	N	2,144	2,147	
No. of Pages	Spearman *rho*	0.123	−0.069	0.005
	Sig.	0.000	0.043	0.885
	N	872	872	870

Colon, downloads and citations

There were 325 titles with and 1,847 titles without colon. Statistics showed that titles with colon tend to be longer ($Mdn = 10$) than titles without colon ($Mdn = 9$; $U = 222{,}427$, $P < 0.001$; $r = 0.16$). Titles with colon also received fewer citations ($Mdn = 9$) compared to titles without colon ($Mdn = 12$; $U = 267{,}253$, $P = 0.012$; $r = 0.054$) Table 6.

Table 6 Man-Whitney *U* tests results for presence of colon paired with other variables

	Median		*U*	*P*	*r*
	Without colon	With colon			
Title length	9	10	2,22,427	0.000	0.16
Times cited	12	9	2,67,253	0.012	0.047
Downloads	2,705	2,777	2,83,916	0.120	0.033

Discussion

The findings showed that articles with question-type titles are downloaded more but cited less compared to articles with descriptive or declarative titles. As research by Ball (2009) showed, authors mainly use question marks in article titles because of marketing as they think questions appeal to the curiosity of users and attract them. This obviously means more downloads but not every download results in a citation. Download can be done by any one, undergraduate students who want to do their assignments or a patient who desperately looks for more information on his or her disease. But citations are only done by authors. Downloads do not necessarily result in using or even reading the article. But citation actually means that the article is somehow related to what the citing author writes and that it has to pass his or her quality control. Therefore, attractiveness of title is not sufficient for an article to be cited, but it might be enough for it to be downloaded.

The data also showed that declarative titles are both downloaded and cited less than descriptive titles though the difference in our study was not statistically very significant. However, this might be a little unexpected as declarative titles are normally those with a clear findings and in that regard more attractive (as well as informative). On the one hand one can argue that ambiguity of title up to a certain point can lead to more curiosity of users and therefore, it may result in more downloads. When users, based on the title, consider an article potentially relevant to their information needs but cannot obtain enough information from title or abstract, they might decide to download the article and skim through it. Also if the title is crystal clear and declarative, then users might get what they need and see no need to consult the full-text of the article.

Our findings lend support to those of Yitzhaki (1994, 2002) in that they show articles with longer titles are more likely to be longer in length. This might be because longer articles are normally the results of large projects and the main medium for publishing the projects' outcomes and therefore they are longer in length. Unlike the study by Jacques and Sebire (2009) that revealed a positive correlation between title length, presence of colon, and citations, our study showed a negative correlation (though a weak one) between title length, presence of colon, and downloads and citations. It also turned out that titles with colon tend to be longer than titles without colon. In the literature (see e.g., Moore (2010)) it has been argued that the advantage of longer titles and titles with colon might be because they include more keywords and therefore are more findable and visible in databases. In titles that have colons especially in scientific fields (not humanities) what comes before colon is normally main keywords. In article writing it is recommended to start the title with keywords if possible for it will make it easier to be noticed by users while they are scanning through their search results. However, in today's databases with the ability of full-text indexing the above assumption may not be necessarily true.

The positive correlation (0.56) we found between downloads and citations is way greater than the correlation found by Moed (2005) (0.22) or by Hitchcock et al. (2002)

(0.27). This might be because the journals in our study were all prestigious *open access* journals.

It is essential to make clear that all findings we presented in this article relate to six open access journals in the field of life and medical sciences and findings may not be correct for other fields. Our dataset was not very large, and although differences and correlations were found in our data, the effect size values (r) were small in most of cases.

Conclusion

This is probably the first study to investigate the relation of title type and downloads as well as citations. None of the past studies studied the three types of titles and they all also focused only on citation data. The main conclusions that can be drawn from the results and discussions above are as follows:

- There are differences between articles with different types of titles in terms of downloads and citations, especially articles with question titles tend to be downloaded more but cited less than the others.
- Articles with longer titles are downloaded slightly less than the articles with shorter titles.
- Titles with colon tend to be longer and receive fewer downloads and citations.
- As expected, number of downloads and citations are positively correlated.
- Overall, title characteristics have more impact on the number of downloads than number of citations.
- Syntax and content analysis of titles as well as qualitative in-depth studies of users' attitude are required to uncover the reasons and the mechanism of the impact of title types on the use of articles. Similar studies could also be fruitful in exploring the possible impact of other characteristics of an article on its marketing and use.

References

Anthony, L. (2001). Characteristic features of research article titles in computer science. *IEEE Transactions on Professional Communication, 44*(3), 187–194.

Ball, R. (2009). Scholarly communication in transition: The use of question marks in the titles of scientific articles in medicine, life sciences and physics 1966–2005. *Scientometrics, 79*(3), 667–679.

Buxton, A. B., & Meadows, A. J. (1977). The variation in the information content of titles of research papers with time and discipline. *Journal of Documentation, 33*(1), 46–52.

Diodato, V. (1982). The occurence of title words in parts of research papers: Variation among disciplines. *Journal of Documentation, 38*(3), 192–206. doi:10.1108/eb026728.

Gustavii, B. (2008). *How to write and illustrate scientific papers* (2nd ed.). Cambridge: Cambridge University Press.

Hartley, J. (2007a). Colonic titles. *Journal of the European Medical Writers Association, 16*(4), 147–149.

Hartley, J. (2007b). Planning that title: Practices and preferences for titles with colons in academic articles. *Library and Information Science Research, 29*(4), 553–568.

Hitchcock, S., Brody, T., Gutteridge, C., Carr, L., Hall, W., Harnad, S., et al. (2002). Open citation linking: The way forward. *D-Lib Magazine*, *8*(10). www.dlib.org/dlib/october02/hitchcock/10hitchcock.html. Accessed 5 April 2011.

Jacques, T. S., & Sebire, N. J. (2009). The impact of article titles on citation hits: An analysis of general and specialist medical journals. *Journal of the Royal Society of Medicine Short Reports, 1*(2), 1–5.

Kane, T. S. (2000). *Oxford Essential Guide to writing*. New York: Berkley.

Lewison, G., & Hartley, J. (2005). What's in a title? Numbers of words and the presence of colons. *Scientometrics, 63*(2), 341–356.

Manten, A. A., & Greenhalgh, J. F. D. (1977). Titles of scientific papers. *Animal Feed Science and Technology, 2*(1), 1–6.

Moed, H. F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. *Journal of the American Society for Information Science and Technology, 56*(10), 1088–1097.

Moore, A. (2010). *Do Article Title Attributes Influence Citations*?. New York: Wiley-Blackwell Publishing News.

Sagi, I., & Yechiam, E. (2008). Amusing titles in scientific journals and article citation. *Journal of Information Science, 34*(5), 680–687.

Soler, V. (2007). Writing titles in science: An exploratory study. *English for Specific Purposes, 26*(1), 90–102.

Tocatlian, J. J. (1970). Are titles of chemical papers becoming more informative? *Journal of the American Society for Information Science, 21*(5), 345–350. doi:10.1002/asi.4630210506.

Wang, Y., & Bai, Y. (2007). A corpus-based syntactic study of medical research article titles. *System, 35*, 388–399. doi:10.1016/j.system.2007.01.005.

Wu, S. F., & Chen, H. J. (2000). Two-stage multiple comparisons with the average for normal distributions under heteroscedasticity. *Computational Statistics & Data Analysis, 33*(2), 201–213.

Yitzhaki, M. (1994). Relation of title length of journal articles to number of authors. *Scientometrics, 30*(1), 321–332.

Yitzhaki, M. (1997). Variations in informativity of titles of research papers in selected humanities journals: A comparative study. *Scientometrics, 38*(2), 219–229.

Yitzhaki, M. (2002). Relation of the title length of a journal article to the length of the article. *Scientometrics, 54*(3), 435–447.