March 2018

Examining the Problems and Inconsistencies in the interpolation of English Transliterated names of Persian Language Researchers in Citation Databases

Azar Mohammadzadeh sarab
Department of Medical Library and Information Science, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Maryam Kazerani (corresponding author)
Department of Medical Library and Information Science, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Maryam Shekofteh
Department of Medical Library and Information Science, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Examining the Problems and Inconsistencies in the interpolation of English Transliterated names of Persian Language Researchers in Citation Databases

Azar Mohammadzadeh Sarab

Maryam Kazerani (corresponding author)

Maryam Shekofteh

Department of Medical Library and Information Science
Faculty of Paramedical Sciences
Shahid Beheshti University of Medical Sciences
Tehran, Iran

Abstract

English Transliterated names of Non-Roman language researchers have been indexed in citation databases in various ways and do not follow a specific rule. For this reason, all the works of a specific writer are not retrieved while searching. This problem is also evident in the transliterating the names of Persian language researchers widely. This study has examined the problems and inconsistencies in the interpolation of English Transliterated names of 1301 faculty members of Shahid Beheshti University of Medical Sciences (SBMU)were indexed in Scopus and Islamic World Science Citation Center (ISC) citation databases The results showed that 193 (15%) faculty members have not had indexed scientific production in both databases and 1108 (85%) people have been indexed in one of two databases of their papers. 357(32.2%) have registered their names in more than 2 forms, and 413(37.3%) in 2 forms, and only 338(30.5%) of faculty members have registered their names in one form. Therefore, almost 70% of faculty members have not registered their names in a single form. The compilation of a list of names document based on the frequency of written form in valid databases is a solution that has been proposed to resolve this problem.

Keywords: Name authority control, Transliteration, Citation Databases, Non-Roman Language Researchers, Faculty Members, Shahid Beheshti University of Medical Sciences, Iran

Introduction

The retrieval and organizing the information resources is in relation with the language and hand writing that have been among the primary means of expressing information. The language of most of information databases is English. Therefore, if the non-Roman language name of the researchers is searched on these bases, the use of the form of the Transliteration of the names will be necessary. Studies show that the searching for names in non- Roman languages, such as Japanese, Chinese, Korean, Persian, etc., has many problems $(1-5)$. There are no unique Transliteration rules in Iran to interpolate the names of authors, and organizations and, researchers transliterate their named according to their taste (6), although the ALA Romanization rules are not applicable because of the large difference that have with the common form. Therefore, The names of Iranian writers in foreign sources have not been recorded in the same and fixed way, and there are considerable variations and differences in the transliteration of names from Persian to English, which has been led to many problems in information retrieval (5) since the author's name is an important retrieval access point in most databases; the existence of a distinct and unique form of name for each author is especially important to bring together all his works and to eliminate the problems related to the nominal similarity of individuals.

Two major projects with the names NACO and VIAF can be noted among the popular projects in the field of name authority control in the international level$(7,8)$. But at present, the sharing of the documentations of the National Library of Iran has not been carried out in these projects, and due to this fact there is no authority control of Iranian writers in these projects. The variety of interpolation of English transliterated of Persian language names has wider dimensions about the faculty members of the universities and reduces the recall & precision of retrieved documents and so, one of the main access point of documents, its mean "authors name "affected by this problem. The multiplicity of interdisciplinary research and scholars' unconsciousness in interdisciplinary collaboration and naming

similarities makes hard this problem.(9) Since the H index is one of the criteria for assessing the researchers, if all the works of a researcher are not accompanied by unique form of the name in whole way, it will not be possible to accurately evaluate this index, and the exact number of scientific outputs of researcher will not be determined. There is a need for a tool to solve this problem that integrate and unify the English names of non-Roman language researchers based on accepted criteria. However, in projects like ORCID, it has been tried to solve the problem of naming similarity by assigning specific code to researchers.(10) or in ISIN, which is a naming method for media creator or producers, such as books, television programs, newspapers, papers, and the like, and has been emphasized the unifying the names of producers (11). Another example is the Researcher ID, which is an identification system for scientific authors and writers. This system has been introduced in 2008 by Thomson Reuters. Its objective has been explained to solve the problem of identifying the author and writer. (12)

But none of these plans and designs does respond to multiple Transliterations of a particular name in non-Roman languages. Of course, it's worth noting that in order to join these global patterns, there is also a need for basic tools to make the authors themselves aware of their common name, which have been registered in various forms in various databases. This research examines the diversity of written form of the faculty member's names of Shahid Beheshti University of Medical Sciences in two widely used databases in Iran, means Scopus and ICS, and a solution to the problem has been proposed. If the names of the authors and writers to be recorded uniformly in the databases, the comprehensiveness of the retrieval will increase and the visibility rate of scientific research will increase.

Methodology

3

This survey is a descriptive study and a kind of applied / developmental study. The research population consisted of 1301 faculty members of Shahid Beheshti University of Medical Sciences and Health Services who were employed in the academic year of 2015-2016, that at least one of their papers to be indexed at the Scopus and ICS Citation Databases. Data collection by searching the English names of faculty members of Shahid Beheshti University of Medical Sciences was first collected in Scopus database .If the faculty member had no article indexed in Scopus, his name was searched on the ICS database .In the studies data bases (Scopus & ISC), Various forms of the name of a researcher was identified.

Then, in order to provide a list of unique names , the most common name of the author was chosen as the selected name, and other forms of names were listed below and in alphabetical order and was referred with the sign of the * to selected name.

Findings

Table 1: Frequency distribution of faculty members of Shahid Beheshti University of Medical Sciences among covered centers

faculty members	faculties		Educational hospitals		Research centers		total sum	
	Number	percentage	Number	percentage	Number	percentage	Number	percentage
frequency	547	42	680	52	74	6	1301	100

Table 1 shows that Shahid Beheshti University of Medical Sciences has 1301 faculty members, 547 (42%) faculty members of this university are employed in faculties and 680 (52%) in 14 hospitals and 74 (6%) in Research Centers affiliated with the University.

4

Table 2: The indexed and UN indexed names frequency of faculty members of Shahid Beheshti University of Medical Sciences in the studied data bases

faculty members	Frequency of indexed names in citation database				Frequency of un-indexed names in citation database		total	
	Scopus		ICS		Scopus and ICS			
	Number	percentage	Number	percentage	Number	percentage	Number	percentage
faculties	425	79	26	5	86	16	547	100
hospitals	577	85	10	1	93	14	680	100
Research centers	58	78	2	3	14	19	74	100
Total frequency	1070	82	38	3	193	15	1301	100

Table 2 data shows that the names (10%) of the 1070 faculty members of the Shahid Beheshti University of Medical Sciences have been indexed in Scopus base and names of 3% of the faculty members have been indexed in ICS. (15%) 193 faculty members have not had indexed scientific production at any of the bases.

Table 3: The Frequency of Written Diversity in the Names of Faculty Members of Shahid Beheshti University of Medical Sciences in the studied data bases

Faculty Members	Writing form diversity of the faculty member's names						total	
	One form		Two forms		More than 2 forms			
	Number	percentage	Number	percentage	Number	percentage	Number	percentage
faculties	161	35	191	41	109	24	461	100
hospitals	168	29	196	33	223	38	587	100
Research centers	9	15	26	43	25	42	60	100
Total frequency	338	30.5	413	37.3	357	32.2	1108	100

Table 3 shows that out of 1108 faculty members, only 338 (30.5%) have registered their names in one form in databases. The most diverse form of writing form of the names is related to faculty members of research centers.

5

Discussion

There are serious problems for Transliteration and paraphrasing the Persian alphabet into English, due to the distinct difference between the signs of the Persian alphabet with the signs of the English alphabet. For example, sometimes a phoneme is written with a few signs, and sometimes there is only one written sign for some phonemes. In some cases, the written signs do not represent any phoneme. Therefore, the Transliteration and paraphrasing of Iranian names is encountered with disparity.(5) As the name of Fereidoun Azizi which is one of the most prolific Iranian scientists, has been indexed into four forms in Scopus:

Since the author's name in all databases is one of the most important search options, and it is important to search for resources and find scientific works, and on the other hand, because of the importance of recording the unique name of the researcher in quantitative scientific assessment studies to evaluate a writer based on the number of works or the amount of citation to the works and the calculation of indexes such as the H index, etc. It is necessary to think about authority control and unifying of the names in non-Roman languages.

The findings of this study showed that out of 1301 faculty members of Shahid Beheshti University of Medical Sciences, 357 persons have registered their names in more than 2 forms and 413 in 2 forms and only 338 faculty members have registered their names in one form. . Therefore, according to statistics, almost 70% of faculty members have registered their names in more than one form. These findings suggest that the diversity of written form in the names of Iranian writers and authors that is consistent with Khosravi's research that has examined the diversity of written form of Iranian researchers in ISI(6), and in both studies it has been pointed out to this important point that the registering the Iranian names in

6

outside resources is diverse. Also, the findings of the researches that have been conducted on the authors and writers of the Korea, Japan and China show a diversity of written forms between the authors and writers of these countries. $(1–4,13–15)$

It is clear that the lack of a single instruction for the Transliteration and paraphrasing has been caused to existence of these varieties in the writing and registering the written form of names. On the other hand, it cannot be expected that the user to have all possible forms of writing the author's name in mind, and somehow it is necessary to be referral network from different forms of the name.

In this study, after examining and searching the names of university faculty members at citation databases, the most frequent written form of English names was retrieved and selected as the selected name and referred to it from rest of the forms:

The use of this list of documents can create a coherence and uniformity in the registration of individuals in resources and databases and improve the comprehensiveness and retention of resources. However, this list should be constantly updated, and a group of library specialists try to review and update this list.

References:

1. Chan K, Hu L, Lo P. A Collaborative Project on Chinese Name Authority Control: The HKCAN Model. J East Asian Libr [Internet]. 2000 Feb 1;2000(120). Available from: http://scholarsarchive.byu.edu/jeal/vol2000/iss120/3

2. Kimura M. A modification of the FRAD model for personal names in non-Roman languages. J Doc. 2015 Sep 14;71(5):938–56.

3. Kim S, Cho S. Characteristics of Korean personal names. J Assoc Inf Sci Technol. 2013;64(1):86–95.

4. Kim S. Romanization in Cataloging of Korean Materials. Cat Classif Q. 2006 Dec 8;43(2):53–76.

5. Molavi F. Main issues in cataloging persian language materials in north america. Cat Classif Q. 2006 Dec 8;43(2):77–82.

6. Khosravi M. Study of Spelling Negligence of Iranian Authors&92;'Names in &92. Iran J Inf Process Manag. 2012;27(2):605–661.

7. Loesch MF. VIAF (The Virtual International Authority File) – http://viaf.org. Tech Serv Q. 2011 Feb 28;28(2):255–6.

8. NACO - Name Authority Cooperative Program of the PCC - Program for Cooperative Cataloging (Library of Congress) [Internet]. [cited 2017 Nov 18]. Available from: https://www.loc.gov/aba/pcc/naco/

9. Vanasupa L, McCormick KE, Stefanco CJ, Herter RJ, McDonald M. Challenges in Transdisciplinary, Integrated Projects: Reflections on the Case of Faculty Members' Failure to Collaborate. Innov High Educ. 2012 Jun 1;37(3):171–84.

10. Shields D. What is ORCID? Obstet Gynecol [Internet]. 2014 May 16 [cited 2017 Dec 5]; Available from: http://journals.lww.com/greenjournal/blog/FAQ/Pages/post.aspx?PostID=88

11. MacEwan A, Angjeli A, Gatenby J. The International Standard Name Identifier (ISNI): The Evolving Future of Name Authority Control. Cat Classif Q. 2013 Jan 1;51(1–3):55–71.

12. Scientific T. ResearcherID. com [online]. c2008 [cit. 2008-10-30]. What Res. 2008;

13. Kudo Y. A Study of Romanization Practice for Japanese Language Titles in OCLC WorldCat Records. Cat Classif Q. 2010 Mar 31;48(4):279–302.

14. Agenbroad JE. Romanization Is Not Enough. Cat Classif Q. 2006 Jun 5;42(2):21–34.

15. El-Sherbini M, Chen S. An Assessment of the Need to Provide Non-Roman Subject Access to the Library Online Catalog. Cat Classif Q. 2011 Aug 1;49(6):457–83.