Inconsistencies in the retrieval of Indic names in Web of Science: a case study of women authors

Bhaskar Mukherjee[a] and Subrata Gangopadhyay[b]

[a] Professor & Principal Investigator (NSTMIS-DST Project), Department of Library & Information Science,
Banaras Hindu University, Varanasi – 221005, Uttar Pradesh, India,
Email: mukherjee.bhaskar@gmail.com

[b] Research Scholar & Professional Assistant (Central Library), Department of Library & Information Science,
Banaras Hindu University, Varanasi – 221005, Uttar Pradesh, India,
Email: ganguly.subrata@gmail.com

The present study is an attempt to show the variations in search results retrieved for Indic women names in the Web of Science database. A sample of almost 50 women names was identified from official websites of scientific organizations of India. The names were explored through the 'Basic' as well as 'Author' Search field of Web of Science. The results show a wide variation in search results for Author search and Basic search. Several cases were observed where two author records on the same name and affiliation with a different set of publications came in search results. The study suggests that for getting accurate results searching through hyperlinked author names available in the bibliographic details of an author's publication led to best results. The study further suggests that while submitting publication, journals may ask the authors to input primary element, secondary element, and auxiliary-element of author's names instead of surname/last name and first name/given-name/forename. Additional Tab asking how the author would like to render his/her name in the article may be added to the journal submission page so that indexing databases can correctly capture records.

Introduction

Historically, the identity of human beings is by their names. In ancient times, people were identified only by a forename. But with increasing urbanization and increasing population, they felt the need for more elements to distinguish their identity from others[1]. The practice of the last name was introduced in the 13th century, and it became an established norm when King Henry VIII ordered birth to be registered under the father's name in the 15th century[2].

Names are a valuable source of information from an indexing point of view. However, a person's name can exhibit many variations in published documents, and users searching for a name may enter a variant form not found in documents and text, or not matching the form indexed in the system. For example, the name "*Maria Goeppert Mayer*" (see https://en.wikipedia.org/wiki/Maria_Goeppert_Mayer) can be searched in multiple ways with different search queries likely to yield different search results.

Another challenging issue with name search is that in a name, a misspell affects the search results. Every word in name and its order is important, e.g., in a name like Swapoora Rani, the term Rani is equally important for the accurate representation of an author's name, as it is the author's last/surname name. If the term Rani is attached with the first part of her name like 'Swarooparani' – which is quite common in Indic women names, then it may represent a different author. However, in the name 'P Shobha Kruparani' if the term 'rani' becomes detached as 'P Shobha Krupa Rani', it won't give correct results, as 'Kruparani' is her correct surname.

India is known for its religious and cultural diversity. The given names of Indian women reveal their gender, marital status, birthplace, nationality, religion, etc., but their last name varies significantly after their marriage; some of them retain their ancestral name even after marriage but some of them use their husband's family's ancestral name or both. The influence of various epic, caste, and religion play a definite role in the life of its people, adopting rituals and choosing names after birth (https://en.wikipedia.org/wiki/Indian_name). No standardized format exists to date which can guide as

to how to identify the last name element from the first name element for any name. The vast variation in the pattern of Indian names differs from place to place or community to community.

Research question

There is no standardized pattern of rendering Indian names. Sometimes it appears that one would be lucky to be able to locate the literature that they are looking for by the name of the researcher. Of course, in the case of authors with only first name and surname, there should be little problem in finding information. Authors who belong to south India do not generally use surname/family name. This also result in inaccuracies.

No database has provision to index names according to the region. Furthermore, a database like Web of Science (WoS) wherein before 2006, the author's given name was frequently stored only in initial, the same author was indexed twice - one with his/her full forename and surname and another with surname and initial forename. The possible reason may be some journals earlier insisted on listing given names in initial only whereas others allowed authors to present their full given name according to the tradition in their country. But this creates ambiguity and inconsistency in the database with the popular Indian family name like Singh, Dwivedi, Rao, etc., with exactly same initial given name for two different authors working on same or different fields/organizations. For example, while searching publications by Debashri Ghosh of Central Glass and Ceramics Research Institute (CGCRI) by Ghosh, D., publications of Debashri Ghosh and two more authors, Debarati Ghosh and Dinabandhu Ghosh, who also belong to CGCRI are retrieved.

WoS has claimed that they have worked extensively to overcome such issues, introduced fully integrated Researcher ID, Author search, Author Record and curation mechanism to provide the global research community with an invaluable index to author information[3]. It is therefore pertinent to track how far such efforts work for authors of the Indian subcontinent. The present study is an attempt to check the consistency and accuracy of WoS data for Indian women names.

Review of literature

Most of the existing literature mainly focuses on the name issue from searching techniques perspectives and provides some theoretical as well as practical solutions. For example, Horng-Jyh; Jin-Cheon, and Soo-Guan[4] discussed two approaches to name modeling, namely natural language processing and information retrieval model, and suggested a hybrid approach using fuzzy name matching. Position-as-attribute and position-transition-likelihood principles are deployed for integrating the advantageous aspects of both approaches.

Cheng[5] in his article pointed to the confusion with Chinese names and revealed that Chinese people who emigrate to or do business with Western countries often at times adopt a Westernized name by simply reversing the 'surname- given-name' order to 'given-name surname' order. Towards a solution, he suggested to spell all the author's surname in capital letters and the first and middle name hyphenated.

In the Indian context, Raveenthiran[1] explains the naming issue with south Indian names and suggested that the online submission system of a journal should follow flexibility suit for the cultural variations in the pattern of the author's name. Instead of using terminologies like First name, Middle Name, and Last name, the journal system should ask for *first-order name, second-order name,* and *third-order name*.

To the best of our knowledge, there is no known study that discusses the problem of name searching in international databases like WoS or Scopus and to that extent by considering actual names of authors for a country like India where a wide diversity in names exists. The present study is an attempt to fill the gap.

Objectives of the study
- To evaluate the search fields to find which field yields exhaustive and accurate results for Indian women names;
- To identify the duplicity or anomalies, if any for the highest recall, by analysing the searched result under various name formats; and
- To suggest the best possible ways to overcome the variations.

Methodology

We identified names of women researchers/ scientists/faculty from the official website of the university/organization/laboratories located in different parts of the country. Wherever, photographs were available, the same served as confirmation. In case, photographs were unavailable, the first name or given name has been used to identify gender. In general women's names in India mostly end with letters like 'ee' [Kiranmay Vs Kiranmayee] or 'a

[Amit Vs Amita, Anil Vs Anila], or 'i'[Parmeshwar vs Parmeshwari].

In case the identification remained inconclusive, annual reports, factsheets, and social network sites were consulted. To remain unbiased, we chose at least five scientists from each zone with varying name elements viz., one name element (i.e., name of the scientists, no last name), two elements (i.e., first name and last name), and three name elements (i.e., name-middle name-last name).

Each name was used to search publications in the Web of Science database. Complete last name along with full first name was applied. In case full the first name was unavailable, the initials as they are available on the official website were used. In most of the cases, official websites of the authors enlist only 'recent publications', maybe of last few years.

For the present study, we searched publications of a scientist in both 'Author Search' and 'Basic Search' fields of Web of Science. Chicago Manual indexed Indian names under family then given names separated by a comma. Author last name and the full-first name were used as search string under author

tag. Needless to mention that search results of any author search display author's name and current affiliation of the author. However, it also shows the name of other organizations that the author had served in her lifetime. By inspecting publications from official websites, both results were included. In the basic search tag, on the other hand, we used 'Author' field as well as the 'Organization-enhanced' field to identify exact publications. The searching was made in November 2020 and searched results were compared with the enlisted publications on the profile page of official websites.

Results

Differences in search results of 'Author Search' and 'Basic Search' by author field

Table 1 explains the variation in search results while searching names (20 randomly selected authors from all regions) in the 'Author Search' field as well as the 'Basic Search' field in Web of Science. The search string, mentioned in the bracket, has been chosen based on the name available on the official website of the author/institution.

Table 1 — Variation of search results in basic and author search

Name and organization	Author search	Basic search	Remark
TanusriSaha-Dasgupta (Saha-Dasgupta, Tanusri), SNBNCBS	225	90	Results from Alternative names: Saha-Dasgupta, Tanusri; Saha-Dasgupta, T; Sahadasgupta, T; Dasgupta, T.; Saha Dasgupta, Tanusri Sahaare included. While clicking for download, out of 225 only 214 articles are downloaded.
Swati Gupta Bhattacharya (Bhattacharya, Swati Gupta), BI	44	24	The author is also searched by Gupta Bhattacharya, Swati with the same 44 publications. If only initial first name and affiliation are used, then 121 records appeared. As per ResearchGate, the author has more than 44 publications in those journals that are indexed in WoS.
Sarika Maitra Bhattacharyya (Bhattacharyya, Sarika Maitra), NCL	39	20	Although the name is correctly written, records under Author search and Basic search have significant differences.
Sumana Das (Das, Sumana) CGCRI	38	1	So many authors record with the same name. In the Basic search if the name is rendered in the same way as it is written in Author Search only 1 record appeared.
Manikyamba C (Manikyamba, C), NGRI	85	88	Although Manikyamba is her given name, this given name is to be placed under the last name field in the author search. Alternative name: Manikyamba, Chakravadhanula. Two authors record of the same affiliations with a different number of records appeared. No. of records in each result differs considerably.
Priya S (Priya, S), NIIST	5	16	149 variants of the same name appeared in the author search, but none belonged to the same affiliation of the desired author. In the Author search, her affiliation is to be mentioned CSIR.

(*Contd.*)

Table 1 — Variation of search results in basic and author search (*Contd.*)

Name and organization	Author search	Basic search	Remark
Asha SK (Asha, SK) NCL	60	60	Although 'Asha' is the given name, the publication of this author is only shown when her given name is placed in the last name tag. Two authors' records of different affiliations appeared.
Manju S (Manju, S) SCTIMST	0	11	Through Author Search no record available, but through Basic search with affiliation 11 records available
Shikha, CMERI	2	1	The author has many publications indexed in WoS, but no record of the same name & affiliation available through Author search.
Vandana (Vandana), NPL	1	24	Seventy-seven authors record of the same name with different affiliations appeared. Although Vandana is a given name, for searching this name is to be placed in the Last name field. The search results in the Author search are not exhaustive, Author has more publications than it searched in Author Search.
Inderpreet Kaur (Kaur, Inderpreet), CSIO	79	96	Eight Authors' records of the same name appeared. Alternative name: Kaur, I
S. *Saravanadevi* (*Saravanadevi*, S), NEERI	7	6	Devi here to be attached with the name. Alternative name: Saravanadevi, Sivanesan
R. Nandini Devi (Devi, R. Nandini), NCL	50	40	Nandini and Devi when combined, no results appeared. Although Devi is not a known Last name but here to be used as the Last name. Alternative names: Devi, RN; Devi, R. N.
B L A Prabhavathi Devi (Devi, B. L. A. Prabhavathi), IICT	50	33	Alternative names: Devi, Bethala L. A. Prabhavathi; Devi, BLAP; Devi, B. L. A.; Prabavathi Devi, Bethala Lakshmi Anu Prabhavathi. Here the author can be searched by Devi, BLAP, or Prabhabatidevi, BLAP.
Archanamoni Das (Das, Archana Moni) NEIST	11	7	1 result appeared under Das, Archanamoni, but 11 results search through the search string Das, Archana Moni. Alternative names: Das, Archana M.; Das, Archana Das, AM
Asha Lalwani (Lalwani, Asha), NEERI	0	0	No record was found in this last name. She is known in her publications as Asha Chelani. 2 author records appeared in the same author, one having WoS ID: WoSRID: Y-4417-2019, another one without any ID.
Manohar Cathrine Sumathi (Sumathi, Manohar Cathrine), NIO	0	0	At least 8 publications are searched by string Manohar, CS under basic search and author search, but no record available under the Sumathi, Manohar Cathrine.
Maria-Judith, BDG, NIO	27	6	No search results are available under the name available on the official website. Her publication is accessible only under Gonsalves, Maria-Judith BD.
Lidia DS Khandeparker, NIO	47	37	She is only searched by Khandeparker, Lidia. If DS is added no results are found. However, another author Khandeparker, Rakhee D S is searched with DS and two author records on the same name and affiliation have appeared.
Joao, Maria Hil--da Das Marcus, NIO	0	9	Although she has publications in this string nothing is searchable. She is searched only by the name Joao, H M.

From the 'Remark' column of Table 2, it is seen that for searching authors 'publications, the names as available on official websites alone will not yield comprehensive search results. Furthermore, it is seen that search results differ considerably for 'Author Search' and 'Basic Search'. Therefore, to get better results it is recommended to execute search both in Author search and Basic search. Women with three-part names (last name, middle name, first name) need to be searched in different combinations of the parts.

Table 2 — Duplication of same author records				
Name as in Website	in Author Search	in Basic Search		Remark
	A	B	C	
Manju Y Krishnan, CDRI	19	14	14	Two author records on same name and affiliation with a different set of 11 publications for this author.
Rishemjit Kaur, CSIO	12	11	15	The same authored name working in Commonwealth Scientific & Industrial Research Organisation (CSIRO) with a different set of publications has searched.
Prabha D. Nair, SCTIMST	55	47	50	Three separate entries of the same author's name and affiliation with different author ID and publications are available.
Vandana Prasad, BSIP	41	09	28	Two different entries of the same author and affiliation appeared. The significant publication appeared in *Science* are not included in search result. That work is available in different entry with Prasad, V. However, no linking between two entries is available.
S. Swarnalatha, CLRI	29	24	26	Three author entries of the same name and same affiliation found. Some appeared as Swarnalatha, S. and some Somasundaram, S. Another author entry of Somasundaram, Swarnalatha with same author ID but a different set of publications also appeared.
Tanusri Saha-Dasgupta, SNBNCBS	223	86	88	Two entries of the same author name are searched. In basic search, almost 24 articles appear under affiliation Indian Association for Cultivation Sciences.
Sumana Chakravarty, IICT	161	39	40	Same affiliation and author name but a different set of results for basic and author search appeared.
Suman Kumari Mishra, NML	57	4	45	Two author records of the same name and affiliation with the different number of publications came. If the initial is used as first-name in the basic search option, then only results appeared.
Arpita Ghosh, NML	8	2	2	Twenty author records of the same name appeared. Her publications are searched by the different affiliation, and almost three author records appeared in that affiliation
Nisha P, NIIST	54	24	44	Two author entries of the same name and affiliation with the different number of publications appeared.
Sandhya SV, NIO	0	1	1	Three authors record of the same name with different affiliations (Central Marine Fisheries Research Institute& NIIST) appeared. None was of desired results. Although the author has publications indexed in WoS, no results appeared under author search.

(*Contd.*)

Name as in Website	in Author Search	in Basic Search		Remark
	A	B	C	
Debashree Ghosh, IACS	51	15	15	Two author entries of the same affiliation but with a different number of records. One with WoS ID, another one without any ID. Basic searches provide incomplete results.
Dipali Devi, IASST	76	19	26	Two entries of the same author and affiliation appeared. Another one consists of 5 publications. Alternative names: Gochhait, Debasis Gochhait, D. Basic search results are completely different than author search results.
Elizabeth Jacob, NIIST	27	3	5	Four authors record the same name, two with the same affiliations but separate entries and the remaining two are from different affiliations. No. of records with the same affiliations differs considerably in different results. Alternative names: Jacob, Sajini Elizabeth; Jacob, E; Jacob, Elizabeth R.; Jacob, Elizabeth A.; Jacob, Elizabeth C.
Sanghamitra Bandyopadhya, IITR	22	22	22	Six author records of the same name, two with the same affiliation, and four with another affiliation (Ind. Stat. Inst.) appeared.

Table 2 — Duplication of same author records (*Contd.*)

A= Complete Last name and First name, B= A + Address of Affiliation (institute name used in WoS abbreviated form), C= A + Organization Enhanced

CDRI- Central Drug Research Institute, Lucknow; CSIO - Central Scientific Instruments Organisation, Chandigarh; SCTIMST- Sree Chitra Tirunal Institute for Medical Sciences & Technology, Thiruvananthapuram; BSIP- Birbal Sahni Institute of Palaeosciences, Lucknow; CLRI- Central Leather Research Institute, Chennai; SNBNCBS- S.N Bose National Centre For Basic Sciences, Kolkata; BI- Bose Institute, Kolkata; IICT- Indian Institute of Chemical Technology, Hyderabad; NML- National Metallurgical Laboratory, Jamshedpur; NIIST- National Institute for Interdisciplinary Science and Technology, Thiruvananthapuram; NIO- National Institute of Oceanography, Goa; IACS- Indian Association for the Cultivation of Science, Kolkata; IASST- Institute of Advanced Study in Science and Technology, Guwahati; IITR- Indian Institute of Toxicology Research, Lucknow

Duplication of records by the same author

As we have seen that 'Author search' provides more exhaustive results than 'Basic search', the next attempt has been made to check how far author search results are complete. For this purpose, another set of fifteen authors have been chosen. The results of the analysis are shown in Table 2, on which the column 'Remark' is self-explanatory. While searching the name through the Author search tab it was observed that a considerable number of authors have been indexed more than one time, despite both the author belong to the same name and organization. Table 2 shows such inconsistencies.

The possible reasons for these inconsistencies are because of switch-over of scientists from one organization to another because of which two entries appeared in the searched results. But in some cases, it is observed that more than one entry of the same author with the same affiliation also appeared in the search results.

Inconsistencies in search results

The next attempt has been made to test how far the search records are accurate and complete for an individual author while searching their publication by the Author search tab. The author search function of WoS only allows to search authors by 'Last name and 'First name' and middle initial(s). The anomalies have been pointed in Table 3. Despite the fact, WoS is one of the leading databases, several cases included the publication of other similar authors with the publication of another author. If the number of publications by an author is more than a hundred then it becomes very difficult to identify such anomalies.

Discussion

From the above set of data, in international databases like WoS, getting a complete publication of a woman author is quite difficult. 'Author search' in most cases provides more exhaustive results but not complete. It may be argued that the significant difference in search results between 'author search' and 'basic search' is because in author search all publications of an author that belong to her current as well as earlier affiliation are included however, in basic search we restrict the results by adding current organization name, where the author currently

Table 3 — Inaccuracies in total publications by an author
(based on Author's Last name and First name in Author Search)

Name	Records	Anomalies
Divya Singh, CDRI	80	Thirty-five authors record of the same name with different affiliations appeared – Alternative names: Singh, Divya Pratap; Singh, Divya Jyoti, Singh, D. At least 5 authors record belongs to our source author.
Pooja Devi, CSIO	70	One entry with 70 publications (ID: K-4199-2015) another with 8 publications. Some of the publications that are searched through WoS are not seen on the official website of the author. E.g., In 2020 author mentioned 14 publications in her credit on the official website but in WoShas indexed 18 publications.
Nandini Das, CGCRI	51	Almost 20% of her total articles have been indexed in WoS that has been written by the different author of the same name belonging to different affiliation.
Charu Sharma, IMTECH	95	Her name with exact affiliation is not searched through the Author search tab but she is searchable in Basic search through author and affiliation. In this tab her 8 publications are retrieved, however, on clicking the Author name in her 8 publications, a further 95 publications appeared under her credit.
Neelam Kumar, CSIO	100	Her name can be searched in the Author search tab, but the search results show publication under her credit which she probably has not authored. On cross-checking of the journal, it observed that some title belongs to optics, other are physics and some are of economics.
Sumana Chakraborty, IICB	164	In our first phase of searching, only 64 records were showing under her credit, which later increased to 164 records. On minute inspection of publications under this name with 164 publications, it was observed that publication that originates from a similar name with a different affiliation is also included here.
R Pratibha, RRI	57	One entry of the same author linked with 57 publications (ID: E-5101-2012) and another entry linked with 9 publications.
Aruna Dhathathreyan, CLRI	30	One entry of the same author linked with 79 publications (ID: AAK-2473-2020). But in the Basic search, actual 30 publications appeared.

CDRI- Central Drug Research Institute, Lucknow; CSIO- Central Scientific Instruments Organisation, Chandigarh; CGCRI- Central Glass and Ceramic Research Institute, Kolkata; IMTECH- Institute of Microbial Technology, Chandigarh; CSIO- Central Scientific Instruments Organisation, Chandigarh; IICB- Indian Institute of Chemical Biology, Kolkata; RRI-Raman Research Institute, Bangalore; CLRI- Central Leather Research Institute, Chennai

working, under 'organization-enhanced' tab. Therefore, the results were only those records that fulfil both the two conditions. However, in some cases, it is observed that the results from basic search showed a larger number of publications by an author than author search, reasons of which is unexplored. Furthermore, in some cases, author searches do not show any results even though the author has a publication of the same affiliation.

The reasons were not clear to us. Therefore, the accuracy and authenticity of the retrieved data cannot be solely left to click of the button. The researchers need to eliminate the noise and spurious records that get crept into the downloaded records because of various factors including the different styles and orders of writing the names and affiliations. Women with similar names might be working in different areas in different institutions. Thus, it would be naïve to rely on the simple downloads without going into

cleaning the data and this problem arises because of homonyms.

During our exploration, we observed that for getting better search results of an author it is essential first to identify at least a few titles written by that author. Of these identified titles, any title may be placed in the basic search tab to get complete bibliographic detail of that title. In WoS, while displaying the search result of any title, it shows the name of all contributing authors is hyperlinked, as it is appeared in the title, the author's full first name in the bracket, author's current and earlier affiliations, and alternative names of that author. By clicking the hyperlink of the desired author name, it is possible to track the total publications of that author in his/her lifetime. However, the greatest challenge by this method is to identify any correct publication by an author as in several cases it was observed that the official website of the authors does not contain any

information about his/her publications. In that case use of social network sites like ResearchGate, Google Scholar may be explored to know the publications of that author.

The Classified Catalogue Code[6]in its rules JA& HD has discussed the rules for handling Name-of-Person. In CCC, the family name has been considered as the primary element and the given name has been considered as a secondary element. Other parts of any personal name have been considered as an auxiliary element. Accordingly, journal publishers while asking authors to submit their articles may also ask them to indicate the primary element, secondary element, and auxiliary-element of their name instead of forename or surname.

Another possible solution to this problem is that the author may be asked to write her name as they are willing to render in their publication. Furthermore, authors should always be asked to render their given names in full instead of abbreviation. It will further minimise the overflow or super-flow of one author's publication with another similar given name. An author may also be asked to fill in the name of other organizations she served earlier to establish the connection between old and new publications.

Conclusion

Although the current study deals with a small sample of Indian women authors, we believe our results may seem true with a large volume of data too. It is a fact that throughout the world, including China, Europe, Africa, South America, no consistent rules exist for rendering author names. The name confusion among Indians is prevalent more with south Indian authors, as their names consist of the father's given name, place of origin, and caste name instead of the surname. Therefore, journal editors & publishers, indexers should understand the cultural variation of the name in various regions and accordingly ask authors to render their names in an unambiguous format. At the same time, the official website of the author must render their author's name in a specific order so that understanding the last name does not become an issue.

Our results show that in WoS the search results for Indian women's names are not always accurate. Even though WoS is working to overcome anomalies, a huge number of inaccurate records exists. We observed that irrespective of regions, the Author search leads to more compressive results than the basic search by name. But not necessarily author search always shows the correct number of records by those authors. In several cases, we observed that more than one author entries for the same author with the same affiliation, and each entry consisting of a different number of records. More research studies with different samples need to be conducted.

Acknowledgments

References

1   Raveenthiran V, Insensitivity of editors and indexers regarding the cultural variations of authors' surnames, *Biochemia Medica*, 26(2) (2016) 164–8.
2   McKie D, What's in a Surname? A Journey from Abercrombie to Zwicker. New York: Random House, 2013.
3   Clarivate, Author data, made better together (2019). Available at https://clarivate.com/webofsciencegroup/article/author-data-made-better-together/
4   Horng-Jyh PW, Jin-Cheon N, Soo-Guan CK, A hybrid approach to fuzzy name search incorporating language-based and text-based principles, *Journal of Information Science* 33(1) (2007) 3-19, doi:10.1177/0165551506068146
5   Cheng Tsung O, The continuing confusion in figuring out the surname of a Chinese author – a proposed solution, *The Chinese Journal of Integrated Traditional and Western Medicine*, 18(4) (2012) 243-244.
6   Ranganathan S R, Classified catalogue code with additional rules for Dictionary Catalogue *Code. Madras: Madras Library Association,* 1964.