Identification of Author and Reviewer from Single and Double Blind Paper

Jatinderkumar R. Saini, Nikita R. Sonthalia, Khushbu A. Dodiya

Abstract—Research leads to the development of science and technology and hence it leads to the betterment of humankind also. Journals and Conferences provide a platform to receive large number of research papers for publications and presentations before the expert and peer-level scientific community. In order to assure quality of such papers, they are also sent to reviewers for their comments. In order to maintain good ethical standards, the research papers are sent to reviewers in such a way authors and reviewers do not know each other's identity. This technique is called Double-blind Review Process. It is called Single-blind Review Process, if identity of any one party, generally authors', is disclosed to the other. This paper presents the techniques by which identity of author as well as reviewer could be found even through Double-blind Review process. It is proposed that the characteristics and techniques presented here will help journals and conferences in assuring intentional or un-intentional disclosure of identity revealing information by the either party.

Keywords—Author, Conference, Double Blind Paper, Journal, Reviewer, Single Blind Paper.

I. INTRODUCTION

RESEARCH is formal work undertaken systematically to increase the knowledge. As the name suggests, research is exploration work done on already completed or on-going search. Research paper is written to present some new things to the scientific community. A research paper is like an investigative report. It is a question that is answered through exploration of a problem or an original idea that sheds new light on a topic. When an author writes a research paper, he or she collects information about a specific issue or topic, analyzes the information, describes the methodology and discusses the findings with final conclusion on the topic.

When a research paper is written, information and facts from a variety of resources are used to support the presented ideas. Books, articles, case studies, videos, interviews, and Web sites are some examples of sources that one uses while writing a research paper. Citing these sources of information in one's work is essential because of following reasons:

a) It gives credit to the author of the original work
b) It allows the audience to identify and find the source material in order to learn more about the topic
c) It gives the presented paper more credibility due to supporting arguments from high-quality sources

J.R.Saini is with the Narmada College of Computer Application, Bharuch, Gujarat, India as Associate Professor and I/C Director. He is PhD from Veer Narmad South Gujarat University, Surat, Gujarat, India (Contact No.: +91-9687689708; e-mail: saini_expert@yahoo.com).

Nikita R. Sonthalia and Khushbu A. Dodiya are in MCA Semester V at Narmada College of Computer Application, Bharuch, Gujarat, India.

According to Doshi [6], Publishing a paper in a peer-reviewed, or refereed journal is a long process, which can take anywhere between a couple of months (very fast) to a couple of years (not unusual). Many steps in this process happen 'behind the scenes' for authors. In other words, authors are not involved in them and have only a vague idea of what happens to their manuscript at each stage. His blog briefly outlines the following stages:

a) Author submits a manuscript:

The manuscript begins its long journey when it leaves its author's hands. Often, this journey is electronic. The authors upload their manuscripts as directed by the journal and this takes only a few minutes. Alternatively, the manuscript is in the form of a printed hard copy, which is then mailed to the journal. Such a paper printed copy can take days or even weeks to reach its destination. It is important to remember that a manuscript can only be submitted to one journal at a time.

b) The manuscript is acknowledged:

When manuscripts are uploaded using the Internet, acknowledgement is received instantly. If the manuscript is sent as an attachment to an e-mail message, the acknowledgement can be prompt although not instantaneous. Journals that require submissions in hard-copy form also usually send acknowledgements through e-mails.

c) The manuscript is sent for review:

Assuming that the manuscript is appropriate in terms of the journal's scope and is having suitable length, it is sent to at least two reviewers who can assess the manuscript. Referees are busy scientists who are rarely paid for their review job. They often take a couple of months to review if they are prompt. Generally, the time taken by them is longer and they may be reminded by the journal's office for completion of assigned work. To ensure an impartial review, the author's identity is not revealed to the reviewers. The reviewers may recommend for the publication of the paper. They may also recommend for acceptance of the paper with or without revisions. It is possible for the paper to be completely rejected by the reviewers.

d) Authors are informed of the result of the review:

If the manuscript is rejected, authors may begin once again by submitting it to another journal. If the manuscript is not rejected outright, authors are invited to revise the paper and to respond to any specific questions the reviewers may have asked. Authors may also request the journal to reconsider the decision to reject if they do not agree with the reviewer and can offer a point-by-point rebuttal.

e) The manuscript is revised and submitted again:

The authors can revise the manuscript as advised by reviewers and submit the revised version. Typically, at this stage, authors are advised to prepare the manuscript by closely following the journal's instructions about style and format and to submit the manuscript in electronic form.

f) The manuscript is copy-edited:

The finally accepted manuscript of the research paper is finally copy-edited by the journal and sent to the printer for publication. Once a manuscript is considered acceptable in terms of its substance, it is further processed for spelling, grammar, punctuation, style, and completeness. The copy-edited manuscript is converted into suitably formatted pages after incorporating tables and figures, if any. At this stage, the manuscript looks the same as it would look when printed. It is set in the right font, title, headings, and subheadings follow the journal's style, lines of text, and the space between lines conform to the journal's layout, and so on.

g) Page proofs of the article are sent to the authors:

The correctly formatted paper is then sent to its authors for a final check and to eliminate any errors that may have been introduced during copy-editing and formatting. Authors typically receive the proofs as files in PDF format and are requested to return the corrected proofs in a very short time. The short time ranges from a couple of days to a week.

h) Page proofs are corrected and the paper is published:

As a final step, the errors, if any, marked by the authors are corrected and the whole issue of a journal is printed and copies are sent out to the subscribers. The printed version of a journal contains many such papers and at this stage the manuscript is considered as "published."

The review process provides a framework for reviewing assets where you can approve, retire, or delete assets. Reviewers can review and verify the quality and accuracy of the paper content according to their specific area of expertise. Benefits of review process are improvement in the quality of published papers, filtering of the output of papers to the benefit of readers, and a 'seal of approval' that the published work meets certain standards, in particular for lay readers.

There are two kinds of review processes, viz. Non-blind review process and BLIND review process. A non-blind review is the process where a reviewer and author are known to each other. Some scientists feel this is the best way to prevent malicious comments, stop plagiarism, prevent reviewers from drawing upon their own 'agenda' and encourage open and honest reviewing. On the other hand, the others argue the opposite view of this. They see open review as a less honest process in which politeness or fear of retribution may cause a reviewer to withhold or tone down criticism. For example, junior reviewers may hesitate to criticize more esteemed authors for fear of damaging their prospects. Independent studies tend to support this.

A blind review process consists of another two types. They are known as Single-Blind Review and Double-Blind Review. In Single-blind reviewing, the reviewer is unknown to the author, but the identity of the author is known to the reviewer. Double-blind review is a term used where the identity of an author is concealed from a reviewer. This is done to prevent the case where the knowledge of authorship biases the review process. The current paper uses the terms 'Single Blind Reviewing' and 'Double Blind Reviewing', as well as their letter acronyms SBR and DBR, respectively.

In SBR, the names of the reviewers are hidden from the author. This is the traditional method of reviewing and is, by far, the most common type. The main advantage of this process is the anonymity of the reviewer. This allows for impartial decisions free from influence by the author. The disadvantages of SBR are listed below:

a) Authors fear the risk that reviewers working in the same field may withhold submission of the review in order to delay publication, thereby giving the reviewer the opportunity to publish first.

b) Reviewers may use their anonymity as justification for being unnecessarily critical or harsh when commenting on the author's work.

The DBR process greatly assists in the elimination of any perception of bias for or against an author or institution. The bias based on name, recognition, country, gender, etc. is eliminated in this process. This helps to assure the authors that the submitted papers are judged equally and that the submitted paper and not the author(s) and/or their affiliation(s) are judged. There are many advantages of DBR and they are listed below:

a) DBR avoids a bias towards top-ranked authors and institutions

b) DBR encourages higher-quality reviews

c) DBR increases the probability of non-novel papers getting accepted

However, there are many disadvantages too of the DBR process. These are listed below:

a) DBR gets papers rejected for failing to sufficiently cite the author's own work

b) DBR decreases enthusiasm of reviewers

c) DBR reduces timeliness

d) DBR places extra burden on the authors

In our perspective double blind paper is better than single blind paper because in double blind paper the fairness and quality of paper is maintained. According to Nature [5], the value of peer review is strongly affirmed. They also highlight that 71% researchers have confidence in double-blind peer review and that 56% researchers prefer it to other forms of review. Double-blind reviewing improves the quality of decision making by increasing the focus of the evaluation process on the actual submission, rather than the authors. But there are some ways where double blind review process may not happen in a fair manner. The authors are found to intentionally disclose their identity in paper in a secret way.

II. RELATED LITERATURE REVIEW & METHODOLOGY

Based on our review of related literature, we found that not much work has been done in this specific area. Though, the research has been done on blind papers, review process, citation and data mining on research papers, but research has not been done on identity-finding techniques in the research

papers per se. The research papers that we found during the related literature review are listed in Table I. Table I also lists the comparison of their work with the one proposed through present research paper.

TABLE I
COMPARISON OF CURRENT WORK WITH WORK OF OTHERS

Sr. No.	Author Name	Paper Name	Comparison with Current Work
1	A. Ben Wagner [1]	Author Identification Systems	They have done research on Researcher ID given to author but we have done research on identification of author and reviewer. We find different characteristics to find the identity of authors and reviewers.
2	Hakan J. Holm [2]	Double-Blind in Light of Internet	In this paper author describes the reviewing process and how authors intentionally reveal their identity on internet before paper publication. We, however, have worked on identification of author or reviewer through sources like research paper itself and social networking sites. This is irrespective of whether the authors reveal their identity by themselves or not.
3	Richard T. Snodgrass [3]	Single-Versus Double-Blind Reviewing	In this paper difference between double and single-blind review process has been discussed by the author. We, on the contrary, work towards the identification of author and reviewer. In said paper [3], the author describes the fairness of double blind papers and how it maintains quality and state not to reveal the identity of author. But we proposed techniques using which we can identify the author and reviewer.
4	Shawndra Hill & Foster Provost [4]	The Myth of the Double-Blind Review? Author Identification Using Only Citations	This paper discusses identification of author by citation. They use reference list for identification purpose. They use only one method for identification whereas we have used many features for revealing identification information of author and reviewer.

It was difficult for us to find multiple papers authored by the same person. As a result, it was difficult for us to find a pattern among the papers authored by the same author or set of authors. Another problem faced by us was of finding blind papers. Some authors and reviewers, when interviewed, provided us with many blind papers. These blind papers were used by us for finding the marks of identity of authors in the papers.

We considered many double blind papers in our research work and found the characteristics that preserve the purpose of double blind review process. It is proposed that this will help the journals and conferences to check papers for these characteristics and find out whether author's identity is disclose or not. This could be done before the submission of the paper to the reviewer. We searched many journal sites and many research papers to get those characteristics. We read many papers and got the idea of the process used by authors to disclose their identity. We also conducted many brainstorming sessions with authors and reviewers to get an idea about the topic under consideration. We also interviewed some authors and reviewers to get there point of view on the topic.

III. RESULTS AND FINDINGS

The motive of this research is to see the possibilities as to how the author could be identified by reviewers or how the reviewers could be identified by the authors. It is proposed that the information generated by this research paper will encourage journals and conferences to prevent potential conflicts of interest.

A. Finding Identity of Author

There are many features which help us to identify authors from Double-Blind paper. There is not 100% accuracy but we get information or name of author from such characteristics, as listed below:
a) From reference
b) Giving example of their university name etc.
c) Property of word document
d) Property of PDF
e) Property of excel
f) University name
g) Frequency of word/phrase usage
h) Particular pattern or logo use in document
i) Social networking
j) Personal acknowledgement
k) Citation of submitted work in bibliography
l) Dataset

The descriptions of some of the characteristics to find the identity of author are given below:
• From Reference:

This is also called citation. In this scenario, the authors give the reference of their own research paper or book. Many times, for acknowledgments, the authors also give information about the organization at which they are working which in turn helps to identify them. The URL of locations from where the technical reports or software has been downloaded for use in research work is also often listed by the authors. Such kind of proprietary locations are also helpful in identifying the organization of institute where the author may be working. A look at the most recent paper of the author is a key source of finding marks of identification of this author in his subsequent papers. For instance, if some author has written a research paper on the subject of Artificial Intelligence in June, then there is very high probability that the same author's next paper may be in July of the same year. Further, the chance, that subsequent paper will cite the earlier paper, is also very high.
• Giving Example of their University Name etc.:

In the contents of the paper, an example of the author's university, friend, own name etc. goes a long way in identifying the author. In cases as such, the use of social networking also is very helpful in confirming the author's identify. In one of the papers, we saw that the example of the research laboratory of the institute of the author is used in the

diagram. Because that laboratory is specifically unique and outsiders are not allowed to pursue research there, it was known that the researchers are from a specific institution and department. In one of the papers the author was found to have given his name in the 'test data'.

• Properties of Word Document, PDF and Excel:

In word document, 'Properties' will automatically save the name of Server or the name of the organization of which the computer is used for preparing the manuscript. The last modified date and created date will be there so from that we know the date of research. Even there is also the name of the personal computer or server or user who has last saved it. We saw in many papers' properties and found the name of author as well as their organization. Using social networking also, we can get the marks of author identification. For instance, if we get the organization's name from the 'Properties' and if we know that who is doing research in that organization on a specific, then we can easily identify the authors of the research paper.

• University Name:

If from research paper or from the website publishing the research paper, we can know of the university, then also we can get the name of the authors of the research papers.

• Frequency of Word/Phrase Usage:

Many authors have a habit of using particular word or phrase many times. If the reviewer knows this habit of the author, then the author could be easily recognized. For instance, use of word 'like' and the use of pronouns 'I', 'We', 'She' and 'He', gives an easy clue to find the author's identify based on his stylometric pattern. Even the author's gender could be predicted by this analysis.

• Particular Pattern Use in Document:

Some conference or journals have their particular writing pattern which they give to the authors who submit their paper. The author has to submit their paper in that conference in that particular pattern only. Based on this pattern, it could be easily found out as to which conference or journal, paper was submitted to.

• Social Networking:

Using any of the techniques described so far, if we get the name of university of the author, then it becomes very easy to reach the author's details. If reviewer gets some paper from particular journal or conference, then the information about the author of the paper could be easily made out. This is so because there is very high probability that for such a journal or conference, most of the papers will be either from the host institution or from the institutions nearby. Further, if we get the name of author's friend or colleague from the research paper, then also we can get information about the author. In social network analysis, the task of identity matching is casted as a structural equivalence problem.

• Personal Acknowledgements:

In acknowledgements, the authors mention the name of their university, guide, funding agency and colleagues. From this information, the details and identity of the author of the research paper could be very easily made out. In some papers, we found names of persons from which at-least the location of

author could be easily traced out. For instance, in the acknowledgement section, names like 'Yang' and 'Chang' easily lead us to say that the authors could be from countries like China and Japan. Similarly, the names like 'Balasubramanium', 'Ramesh' and 'Vishvanath' easily lead us to the conclusion that the author could be from India.

• Citation of Submitted Work in Bibliography:

Citation is nothing but a reference or bibliography in which the authors give the referrer research paper reference or cite description by which they get some information or knowledge about the topic or analyze it and give their own review on it. There is subjective evidence that self-citations are an important identifier of paper authorship. Self-citation is most popular feature to find identity of an author. In this case the author gives their own blog or paper's reference.

• From Dataset:

There is some dataset which is available to some author, agency or country only. Based on usage of this dataset, it could be easily found out as to who is actually doing research and hence the author of the research paper could easily be found out. For instance, if the author uses the dataset of some country where the government has not allowed the use of dataset outside the country, then the country of the author could easily be made out.

Table II describes the probability of above characteristics to happen from higher to lower order of likelihood.

TABLE II
AUTHOR IDENTIFICATION CHARACTERISTIC RANKING

Probability No.	Characteristics
1	From reference
2	Citation of submitted work in bibliography
3	Social Networking
4	Property of word document
5	Property of PDF
6	Property of excel
7	Dataset
8	Personal Acknowledgement
9	Giving example of their university name etc.
10	Known university then research topic
11	Particular word use no of time
12	Particular pattern or logo use in document

In characteristics listed in Table II, the most prominent characteristic is the "property of word document" and citation. Our analysis of many papers helped us find that these are prominent sources from where marks of identification of the author could be known. We also found that references and frequency of particular word usage are not much good options to find the author identification. Authors should take care of following things to make the paper completely blind.

a) Eliminate names, contact information, and affiliations from title pages (and anyplace else)
b) Eliminate identity revealing and direct references to funding sources, if possible
c) Eliminate acknowledgments, if possible
d) Scan the text and make sure to use the third person to refer to personal work

e) Make sure figures do not contain any affiliation related identifier
f) De-personalize the work by using anonymous text where necessary
g) Remove excessive self-citations and citations to unpublished work

The following personal identification information about the author could be attempted and easily found from an otherwise claimed to be 'blind paper':

a) Gender of author(s)
b) Number of authors
c) Country/City of author(s)
d) Organization/University/Institute of author(s)
e) Age group of author(s)
f) Marital status of author(s)
g) Religion of author(s)

B. Finding Identity of Reviewer

Finding identity of reviewer from double or single blind papers is done on some criteria which help to find the particular reviewer. Such characteristics used for identifying the reviewer are listed below.

a) Comments
b) Style of reviewing
c) Journal list
d) Social networking

• Comments:

Reviewers have they own method of giving comments. This includes their style of giving same type of comments to various papers.

• Style of Reviewing:

Reviewers have their own style of reviewing the papers. For instance, during interview, some authors told us that their good quality papers were rejected or accepted with modifications while otherwise weak papers were accepted or accepted without modifications. The point here is that reviewers may focus on some particular thing during the review process.

• Journal list:

Many journals have their own set of editors and reviewers. They put the name of such editors and reviewers on the website of respective journal too. It is very easy to predict the name of reviewer from such a list of reviewers published by the journal publishing agency. For instance, imagine a journal or conference having 40 Reviewers, out of which 5 have expertise in Artificial Intelligence. Given a paper of Artificial Intelligence to such a journal or conference, there is a high probability that the paper will go to one of these 5 reviewers only.

• Social networking:

By social networking also, we can get the information about reviewers. For instance, if some paper is sent to a reviewer and it is known that the proposed kind of research work is being done at a particular location only, then using social networking or one's contacts, the author can easily reach the reviewer.

Table III describes the probability of above characteristics to happen sorted in order from higher to lower probability.

TABLE III
REVIEWER IDENTIFICATION CHARACTERISTIC RANKING

Probability No.	Characteristics
1	Journal list
2	Social networking
3	Comments
4	Style of reviewing

The concepts propounded in this paper can be used by journals and conferences to make sure that the double-blind paper review process indeed remains double-blind, irrespective of un-towards intentions of either the author or reviewer or both the stake-holders. This type of system can also be automated.

IV. CONCLUSION

The concepts proposed here help us in identifying the author from a claimed blind paper. The journals and conferences can verify, using the concepts given here, whether the submitted paper is actually blind or not. They can check for knowingly or un-knowingly introduced identification marks in the submitted paper. Similarly, this can be assured that the reviewer also is knowingly or un-knowingly not able to reveal his identity. The number of reviewers is less than the number of authors and so it is easy to find the reviewer's identity by the techniques discussed here.

We can develop a system that takes research paper as an input and uses the characteristics discussed here to find out the identity of author and reviewer. The said system could also be used by conferences and journals to check the paper before submitting it to reviewer. It is proposed that this will stop the intentional or un-intentional disclosure of identity by author as well as reviewer to the other party.

REFERENCES

[1] B. Wagner, Author Identification Systems, *Issues in Science and Technology Librarianship*, 2009.
[2] Hakan J. Holm, Double-Blind in Light of Internet, *Snodgrass-Sigmod-2006*, Available online: http://www.acm.org/pubs/prepub eval.html.
[3] Richard T. Snodgrass, Single - Versus Double-Blind Reviewing, *Editorial,* Available Online: http://www.cs.utexas.edu/users/mckinley/notes/snodgrass-tods-2007.pdf.
[4] S. Hill and F. Provost, The myth of the double-blind review? Author identification using only citations, *ACM, SIGKDD Explorations.*
[5] Editorial, Should there be author anonymity in peer review?, *Nature,* ISSN: 0028-0836, eISSN: 1476-4687, vol. 451, pp. 605-606, doi:10.1038/451605b, Available Online: http://www.nature.com/nature/journal/v451/n7179/full/451605b.html.
[6] Yateendra Doshi, blog entry dated 18 November, 2009; Available Online: http://blog.editage.com/?q=from-draft-to-publication-a-research-papers-journey.