Transgender people are four times as likely to have a household income under $10,000, and twice as likely to be unemployed as the average American.

ADVICE

The Excessive Power of *Ctrl+C* and *Ctrl+V* in CS Research and Career Development

Various types of plagiarism are common in academia. Namely, (i) word-by-word or verbatim plagiarism: turning contents into a carbon copy, (ii) paraphrasing plagiarism: copying the underlying meaning of the content in a "smart" manner by slightly tweaking the tone of sentences, (iii) idea plagiarism: capturing early/unpublished ideas belonging to someone else, and (iv) authorship plagiarism: masquerading as an original author of a different author's work. Moreover, occasionally, reviewers may demand their own papers to be cited (no matter whether the references are relevant or not), leading to a new type of plagiarism that we call "citation plagiarism." In this article, we mention three real plagiarism stories, and show ways to reduce incidents of plagiarism. It is not our intention to blame or shame any one journal, editor, reviewer, university, student, professor, or author. Instead we will give an account of three plagiarism stories to demonstrate how insidious the problem is.

Recently, we were the victims of plagiarism. We found clear evidence of paraphrasing plagiarism in a published paper. It was a random online search that caught our eye. A familiar figure was identical to one published in a paper from a couple of years ago. Curiosity lead us to discover 90 percent of the content of that randomly searched paper was paraphrased from not only our paper, but others (available online) too. We approached the editor-in-chief of the journal of the newly published paper, who found the same issue, leading to a retraction of the paper, after some efforts.

The height of verbatim plagiarism may be illustrated by the following real story. Jure Leskovec, Anand Rajaraman, and Jeffrey D. Ullman wrote a book entitled *Mining of Massive Datasets* (2011). Since the book is available online, someone by the name of Seyed Hossein Ahmadpanah simply copied the entire thing. Seyed only changed two elements of the book: the title and the author's name.[1] Seyed used CreateSpace, Amazon's book-production facility, to produce the hard copies, and made the fake book available on Amazon for $25.

Recently, we were assigned a paper to review for a well-regarded journal. The writing and contents of the paper were extremely poor, which was not an unusual problem until we noticed that the authors' affiliation was listed as MIT. It was unexpected to see MIT authors submitting a paper of that caliber, so we investigated further. We discovered they were not actually enrolled at MIT, or any other U.S. university. Since we were curious to know what went behind this blunder, we did not reject the paper outright, but instead asked for a major revision. The only question that the authors responded to was regarding their affiliation, they claimed they mistakenly wrote in the wrong affiliation.

After reading these stories, a natural question arises in our mind: Why is plagiarism detection getting harder? There might be various possible reasons:

1. **Faulty tools.** Many online plagiarism detectors can only find verbatim plagiarism, not paraphrasing plagiarism. For example, Google Scholar does not count a citation if a paper is not properly cited in a new paper.

2. **Fake reviewers.** In some plots, authors first write a fake paper, and then submit fake reviewers' information that turns out to assign the paper for review to the authors or to their friends.[2]

3. **A lot of papers and few reviewers.** Some top-notch journals/conferences that promise to adhere to the standards usually face a crisis of available reviewers. Most authors wish to submit their papers to good journals; however, the lack of expert reviewers significantly impacts the review process.

4. **Time and effort.** Some students wish to complete their degrees quickly without devoting significant effort. An identical notion is also applicable to some faculty (usually at the beginning of their career), who wish to publish more to quickly earn a promotion. These are the driving factors for plagiarism in terms of manipulating data, algorithms, or system settings.

5. **Trust versus truth.** Journals build up the trust of researchers through impact factors, which are, occasionally, false due to citation plagiarism and self-citation (the number of citations from a journal article to articles published in the same journal). Moreover, numerous papers are submitted to journals for review. Hence, it is hard to accomplish a uniform review process in any journal with a good impact factor.

---

1 See Prof. Jeffrey D. Ullman's May 2017 Google+ post; https://plus.google.com/u/0/110300059908499284579

2 http://retractionwatch.com/2017/04/20/new-record-major-publisher-retracting-100-studies-cancer-journal-fake-peer-reviews/

So what now? Is this the end of good papers or an end of the existing ethical system for the research or review community? Is there any solution to prevent plagiarism? In order to answer these questions, we need to first answer the following questions:

1. What must an editor do? To prevent authorship plagiarism, the editor must first verify authors and reviewers (as suggested by the authors). The editor may also investigate the reviews to some extent, and avoid citation plagiarism by reviewers or him/herself.

2. Single- or double-blind review? The double-blind review process can only help if the paper is not assigned to a fake identity. However, the double-blind review has its own merits and faults, and is a whole topic worthy of discussion. Nevertheless, some leading conferences, like SIGMOD and AIIDE, believe in the double-blind rule to support a fair review system.

3. Why not publish reviews and acknowledge reviewers? This could be an influential step toward an honest review process. After accepting a paper, the journal would publish the reviews with the paper. In addition, rewarding the best reviewers of the year/quarter/volume would also help in improving review skills. Springer's *Distributed Computing* announced the first best reviewer award in 2012.

4. What should be done after identifying a plagiarist? A combustible question is who is responsible for figuring out plagiarism: editors, reviewers, the journal team, or the authors whose paper is plagiarized? In addition, once a journal uncovers an author as a plagiarist, the journal may commit to verifying all the past papers of the plagiarist author in that journal. The chances are high that a plagiarist author is a repeat offender. For example, the author of the fake book in our second story conducted verbatim plagiarism with many books, which are still available online. The journal may take strict actions against the accused one. Although journals should share information, this may open publishers up to liability.

5. How much time? The limited time frame is a major hurdle in achieving an honest review. A possible solution would be to review a paper within a hierarchy of students of the reviewing professor. Of course, during the whole process, reviewers should remain polite and complete a timely review.

Biographies

Nisha Panwar is currently a post-doctoral scholar at University of California, Irvine. She did her Ph.D. within the Department of Computer Science, Ben-Gurion University of the Negev, Israel. She received her M.Tech. in computer engineering from the National Institute of Technology, Kurukshetra, India in 2011. Her research interests include security and privacy issues in IoT such as smart homes, buildings, vehicular networks, computer network and communication, and distributed algorithms.

Shantanu Sharma received his Ph.D. in computer science in 2016 from Ben-Gurion University, Israel, and M.Tech. in computer science from the National Institute of Technology, Kurukshetra, India, in 2011. He was awarded a gold medal for the first position in his M.Tech. degree. Currently, he is pursuing his post doc at the University of California, Irvine assisted by Prof. Sharad Mehrotra. His research interests include designing models for MapReduce computations, data security, distributed algorithms, mobile computing, and wireless communication.