Learned Publishing, 23: 9–14
doi:10.1087/20100103

CASE STUDY

CrossCheck: an effective tool for detecting plagiarism

Helen (Yuehong) ZHANG
Zhejiang University Press

ABSTRACT. The plagiarism detection service CrossCheck has been used since October 2008 as part of the paper reviewing process for the Journal of Zhejiang University – Science (A & B). Between October 2008 and May 2009 662 papers were CrossChecked; 151 of these (around 22.8% of submitted papers) were found to contain apparently unreasonable levels of copying or self-plagiarism, and 25.8% of these cases (39 papers) gave rise to serious suspicions of plagiarism and copyright infringement. Four types of copying or plagiarism were identified, in an attempt to reach a consensus on this type of academic misconduct.

Introduction

CrossCheck (http://www.crossref.org/CrossCheck) is an international project intended to help publishers cope with the increasingly high incidence of plagiarism.[1] CrossCheck helps to protect the original authors' copyrights, and helps to improve authors' behaviour by identifying instances of academic plagiarism. It is led by the Publishers International Linking Association (CrossRef); several global publishing groups are participating.[2] In 2008 CrossCheck won the ALPSP Award for Publishing Innovation.[3]

In October 2008 the *Journal of Zhejiang University – Science (A & B)*,3 which is supported by the National Natural Science Foundation of China, became the first member of CrossCheck in China.[4] CrossCheck is used as part of the journal's review process. Each paper is CrossChecked twice: the first check takes place before it is sent to international reviewers; a second check takes place just before 'online-first' publication, to ensure that no potential plagiarism is missed owing to the inevitable time-lag in updating the CrossCheck database. The date of the latest CrossCheck is included on the first page for each journal paper (Figure 1) for the information of readers, authors, and databases. The majority of authors behave correctly, submitting papers that bear little or no similarity to other published papers. However, around 22.8% of papers appear to contain unreasonable copying or self-plagiarism, and about a quarter of these give rise to serious suspicions of plagiarism and copyright infringement; in some cases, the similarity with the plagiarized original was as high as 83%.

Four distinct types of plagiarism were identified, which we consider sufficiently serious to be considered as a form of academic misconduct:

Helen Zhang

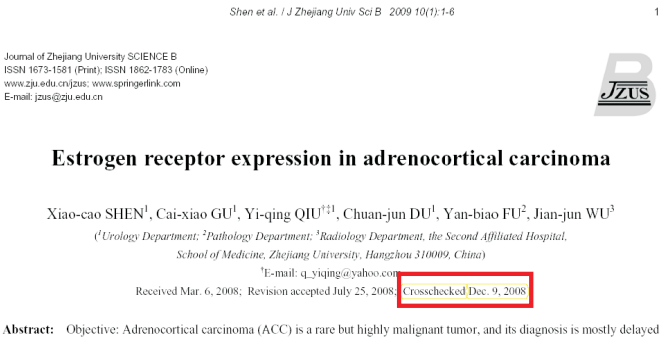

Figure 1. The date of the most recent CrossCheck is indicated on the first page of each journal paper.

1. duplicate publication;
2. self- (or team) plagiarism;
3. direct copying of Methods section, with new data inserted; and
4. uncited or excessive extracts.

A fuller report of the findings, in Chinese, was published in *ScienceTimes*.[5]

around 22.8% of papers appear to contain unreasonable copying or self-plagiarism

Duplicate publication

CrossChecking identified the fact that a few authors had contributed almost identical papers to several journals, or had submitted – completely unchanged – papers previously published in conference proceedings or electronic journals. If the similarity is more than 40–50%, we automatically reject the article on the basis of duplication. For example, in May 2009 the final CrossCheck identified one article which duplicated about 78% of the content from a paper by the same author published in an IEEE journal in early 2009 (Figure 2).

Identification of duplicated text is not difficult using CrossCheck. However, CrossCheck is currently unable to check duplication in figures and tables, so we have recourse to other sources (Google, PubMed Central, etc.) for further analysis of articles highlighted by CrossCheck. Comparison of a paper from France and one from Burkina Faso showed that, while only 18% of the main text was duplicated, the references were identical (Figure 3). When we referred to PubMed Central's full-text database, we discovered that the figures and tables were completely duplicated from the earlier publication, so that the actual duplication was nearer 80%. This shows that our editors cannot rely on CrossCheck alone, but also have to make additional efforts to detect duplication at the time of submission and of publication.

In another typical example, a paper was found to have the same abstract as another previously published paper. Further investigation revealed that the author's Ph.D. thesis had already been published online through a university press, and that the author had also already published, before Ph.D. graduation, two papers containing the

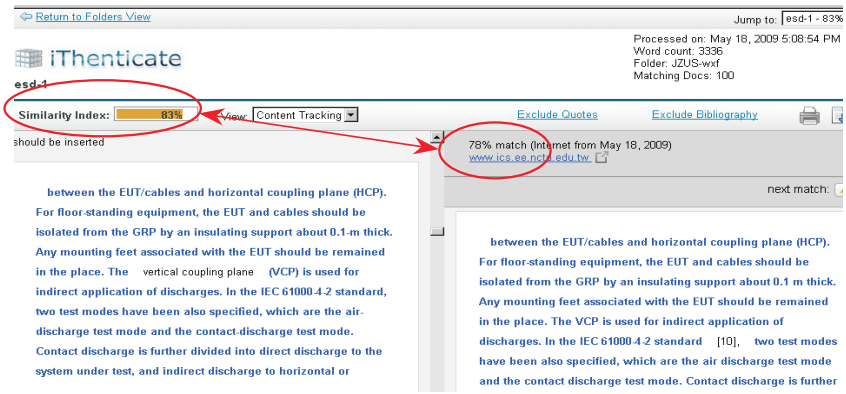

Figure 2. Duplicate publication.

core content of his thesis. However, he argued that although 39% of the paper replicated previous publications, the rest of the material was previously unpublished content from his Ph.D. thesis, and thus that the whole paper should still be published. Since the whole thesis was already available online, and its core content had been published five years previously, we felt that it was unacceptable to republish it unless it contained significant new information, and the article was therefore rejected.

In our view, duplicate publication injures the interests of many journals, wastes publication resources and should be condemned both by academia and the publishing industry. As Errami and Garner state,

> the repeated publication of the same results by those who conducted the research is ethically questionable. It not only artificially inflates an author's publication record but places an undue burden on journal editors and reviewers, and is expressly forbidden by most journal copyright rules.[6]

Further work is needed to define relevant criteria.

Self- (or team) plagiarism

Another familiar phenomenon is self-plagiarism (or plagiarism of the publications of other team members). This can frequently be found in papers of authors from the same research programme (Figure 4). Some authors, or even programme leaders, believe that this is justified by different focuses in the same research project, even when the equipment and methods adopted are the same; thus they do not feel it is unreasonable to duplicate parts of the Introduction, Methods and Discussion sections.

However, in our view, once a paper is published the authors should not recycle any of its content in new papers. Self-plagiarism wastes not only the publication resources of journals but also the time of readers. Instead, authors should simply cite previous studies, giving no more than an overview in their current paper. It is preferable to combine the content of several papers together to form a single high-quality paper, rather than repeating some of the contents to form different papers.

Publishers also object to this practice. As Arnout Jacobs, Vice-President of the Science & Technology Department, Chinese section, at Elsevier, says:

> it is a nuisance for journal editors when researchers publish a series of highly similar papers. Often, these papers could easily be rewritten as one single excellent paper. This happens less often in the USA or Europe, where editors or funding agencies check earlier publications routinely as a reference and authors would be judged negatively for publishing multiple papers with the same topics and replicated contents.[7]

Direct copying of Methods section, with new data inserted

This is a particularly common phenomenon in biomedical papers, where all or part of the Methods section may be copied verbatim, only changing some of the experimental conditions and data (see Figure 5). Some authors feel that it is acceptable to copy all or part of the Methods section from a previously published article, simply inserting their own data.

However, we observe that this type of direct copying is hardly ever found in leading journals such as *Science* and *Nature*. In principle we believe that, although much research refers to or repeats others' successful methods in testing new materials and discussing new results, the authors should use their own language to describe and summarize their methods and ideas.

Uncited or excessive extracts

We found that some authors incorporated extracts from other papers without providing citation details (see Figure 6). In one instance, when we raised the matter with the author, he argued that, since his own view was identical to that of the other author, it was acceptable to use the same words without citation. However, such conduct misleads readers into believing that they are reading the author's own words

he argued that, since his own view was identical to that of the other author, it was acceptable to use the same words

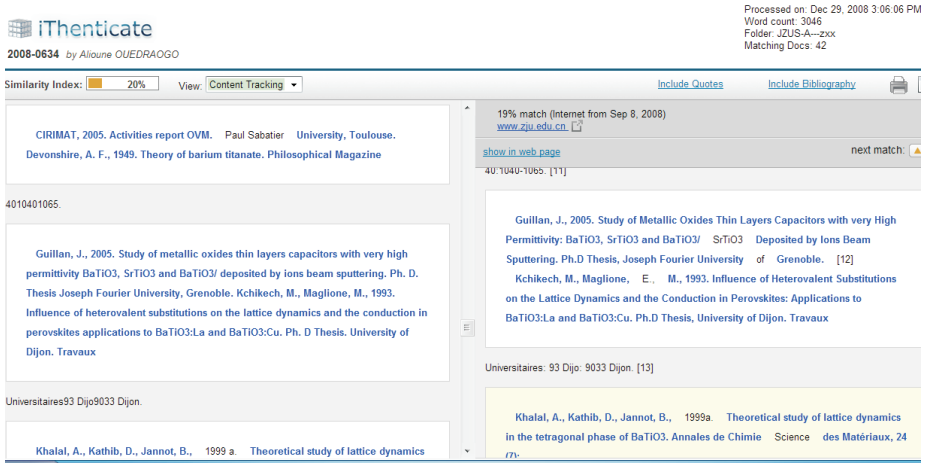

Figure 3. Identical references in two apparently different papers.

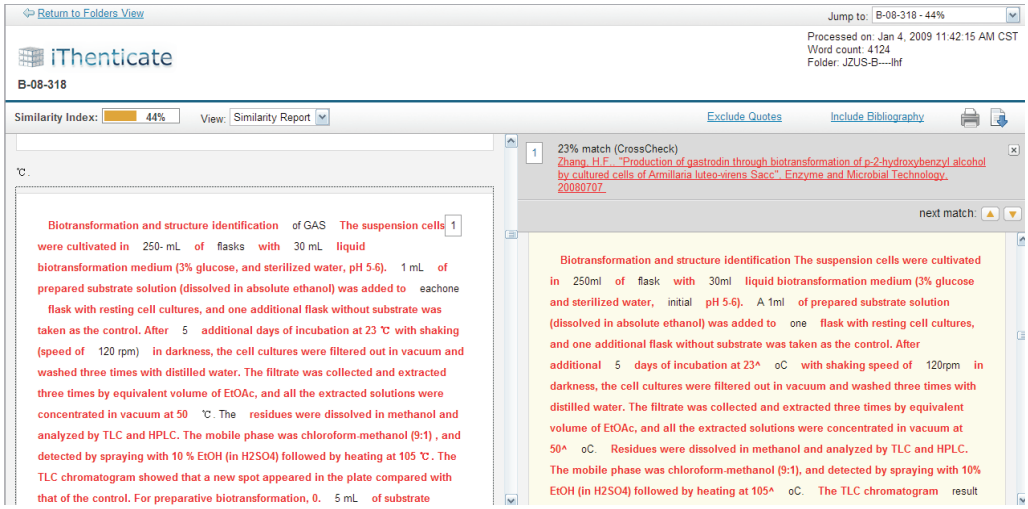

Figure 4. Plagiarism of the work of members of the same research team.

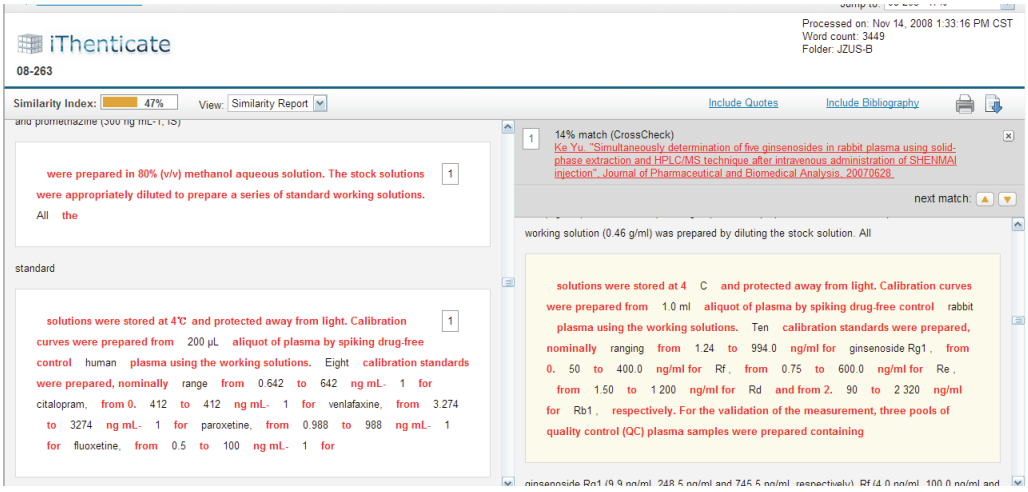

Figure 5. Direct copying of Methods section with new data inserted.

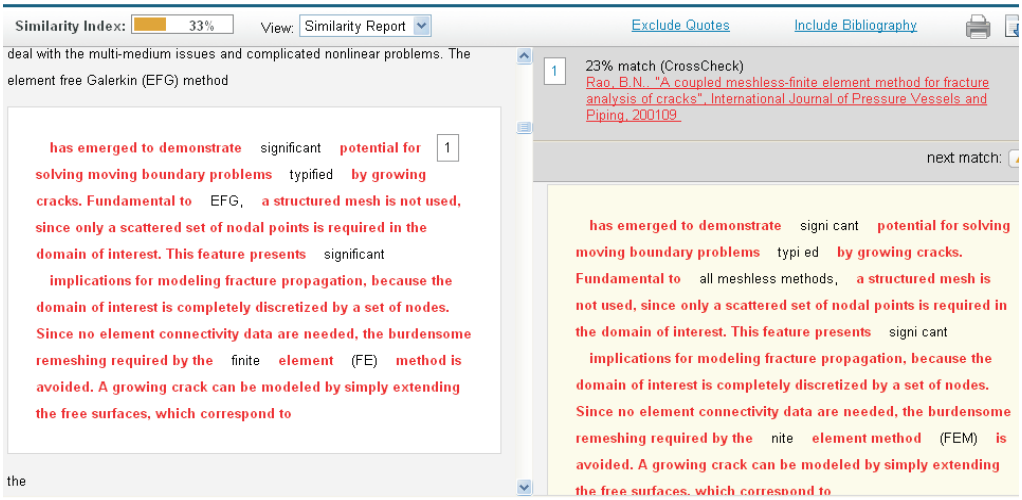

Figure 6. Uncited extracts from other papers.

and, quite apart from its academic impropriety, this is an infringement of copyright.

Sometimes, too, authors believe that, with a full citation, it is reasonable to copy whole paragraphs from other papers; this is not the case, and the 'fair dealing' rules always apply.

The phenomenon of 'copy and paste' is also all too common, particularly in papers from non-English-speaking authors. In a few extreme cases, we found that many sentences and whole paragraphs were identical to those in published papers, and scarcely any of the words were the authors' own (see Figure 7).

The Council of Science Editors gives clear definitions of piracy and plagiarism:[8]

Piracy is defined as the appropriation of ideas, data, or methods from others without adequate permission or acknowledgment. Again, deceit plays a central role in this form of misconduct. The intent of the perpetrator is the untruthful portrayal of the ideas or methods as his or her own.

Plagiarism is a form of piracy that involves the use of text or other items (figures, images, tables) without permission or acknowledgment of the source of these materials. Plagiarism generally involves the use of materials from others, but can apply to researchers' duplication of their own previously published reports without acknowledgment (this is sometimes called self-plagiarism or duplicate publication).

the phenomenon of 'copy and paste' is also all too common

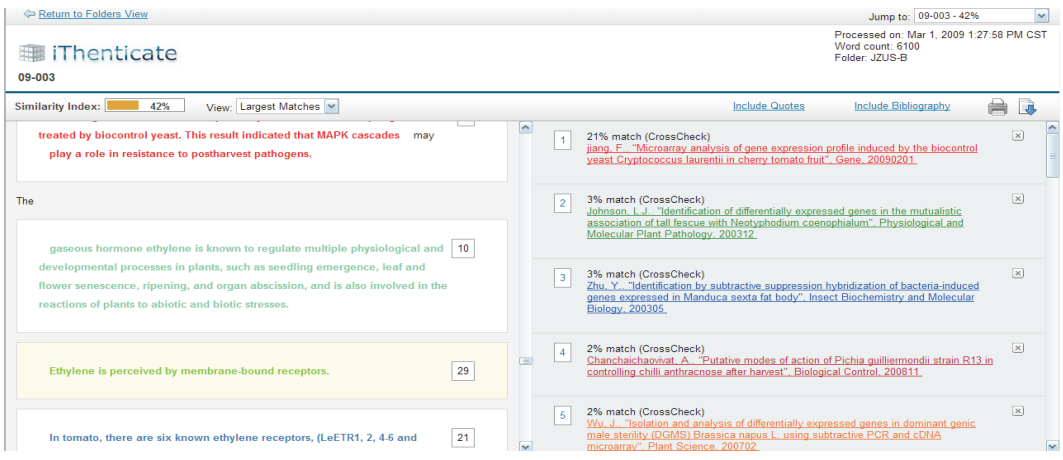

Figure 7. Copy and paste, with almost no original text.

an emphasis on quantity is liable to lead to authors taking short cuts such as plagiarism

Authors should ensure that any article they submit for publication is original and does not contain plagiarized content from either their own or others' work. If an author's text follows the source so closely that the result is more of a the quotation than a paraphrase, it constitutes plagiarism; the author must either completely recast the summary in his or her own words (changing a few words is not sufficient), or quote explicitly.

Conclusions

The importance of science should be measured by the quality of papers rather than their quantity. In China, as elsewhere, researchers and their institutions should be evaluated on the basis of real original research results, rather that on the basis of paper output. An emphasis on quantity rather than quality is liable to lead to authors taking short cuts such as plagiarism.

Academia is not a perfect world; inevitably academic journals all over the world are likely to encounter these or similar problems. As editors, we have a responsibility to promote professional ethics. CrossCheck enables us to see that most scientists do behave ethically. However, it is up to the editorial community to propose criteria and processes for handling these types of academic misconduct. In this way we can help to protect the copyrights of original authors, and promote the healthy development of academic journals.

Acknowledgements

I thank my colleagues Hanfeng Lin, Ziyang Zhai, Xinxin Zhang, and Chunjie Zhang for collecting and analyzing the case study data.

References

1. Meddings K. 2010. Credit where credit's due: plagiarism screening in scholarly publishing. *Learned Publishing*, 23: 5–8. doi:10.1087/20100102
2. http://www.crossref.org/CrossCheck_members.html
3. http://www.alpsp.org/ngen_public/article.asp?id=0&did=0&aid=19899&st=awards&oaid=0
4. English edition at http://www.zju.edu.cn/jzus/
5. Zhang, Y.H. 2009. Academic misconduct as perceived from recent examples. *ScienceTimes*, 7 May (in Chinese), http://www.sciencenet.cn/dz/dznews_photo.aspx?id=6007
6. Errami, M. and Garner, H. 2008. Commentary – a tale of two citations. *Nature*, 451: 397–9. doi:10.1038/451397a
7. Jacobs, A. 2008. How to counter academic dishonesty in STM journals. *Sciencenet*, 5 April. www.sciencenet.cn/htmlnews/2008428182451204881.html (in Chinese)
8. http://www.councilscienceeditors.org/editorial_policies/whitepaper/3-1_misconduct.cfm#3.1.3

Helen (Yuehong) Zhang
Journal Director, Zhejiang University Press
38 Zheda Road, Hangzhou, 310027, China
Email: jzus@zju.edu.cn