

Key issue

CrossCheck



AMY BRAND

Director of Business &
Product Development
CrossRef

CrossCheck: preventing plagiarism in scholarly and professional publications

Anyone who has ever authored a document that involves web-based research understands the temptation to cut and paste bits of text in the course of research and writing. It is simply too easy not to. Of course, most of us use this technique as a loose form of note-taking – collecting quotations, references, and good ideas for the paper we have yet to write.

While ease of copying is only one of many causal factors, it comes as no surprise that plagiarism – passing off someone else's written words as your own – is an increasing problem among students and professionals. A recent article in the journal *Nature* reported on a study that found roughly 1% of the abstracts in the Medline database to be duplicates, after testing a random sample of 62,000 abstracts. The authors assert, "...the duplication of scientific articles has largely been ignored by the gatekeepers of scientific information – the publishers and database curators".¹

CrossCheck is scholarly publishing's response to this charge. Although there are already commercial plagiarism detection tools on the market, they have not been particularly useful for filtering works submitted to academic journals because they have not had access to the relevant full-text literature to check against. CrossCheck is, first and foremost, about creating the needed, comprehensive database of authoritative research content for plagiarism checking purposes. Just as full-text indexing makes a document discoverable through search regardless of whether it is open or paid access content, all

content in the CrossCheck database is functionally checkable regardless of access model.

How it came about

At the 2006 CrossRef Annual Meeting, publishers were asked to rank potential new CrossRef services; creation of a cross-publisher system to verify the originality of submissions emerged as a top priority. In the months that followed, CrossRef staff and Board members reviewed the technical, policy, and business feasibility of creating such a service. The CrossCheck pilot launched in August of 2007 in partnership with plagiarism detection vendor iParadigms and with eight publishers: the Association for Computing Machinery, BMJ Publishing Group Ltd, Elsevier, the Institute of Electrical and Electronics Engineers, IUCr, The New England Journal of Medicine, Taylor & Francis, and Wiley-Blackwell.

What it is and what it isn't

CrossCheck is not itself a plagiarism detection service, but rather a multi-faceted initiative to facilitate the prevention of plagiarism in academic publishing. Its top priority is the creation of a comprehensive database of scholarly and professional content for plagiarism filtering purposes. Plagiarism detection systems employ a technique called text fingerprinting, in which an uploaded document is analyzed and compared against a database of pre-analyzed content. The system is fine-tuned to detect and report degrees of textual overlap between the

target document and any matching documents in the database. It is up to the person reading the report output by the system to gauge how much and what kind of overlap is actually plagiaristic. Clearly, such a system is useless without the relevant database of content, and without trained individuals interpreting the reports.

Other components of the CrossCheck initiative include drafting and negotiating business terms via which published content is accessible for indexing and checking; establishing prerequisites for publisher participation and compliance, along with ways that publishers can promote their participation in the program to deter future plagiarism; and working with manuscript tracking vendors and policy organizations to co-ordinate the development of best practices that will help publishers implement CrossCheck most effectively and ethically.

When publishers opt into the programme, they are offered the opportunity to sign an agreement with a plagiarism detection vendor that has been vetted by a committee of CrossRef members and staff to ensure that publishers' shared concerns about the indexing process, content security, pricing, access, support, etc. are already addressed. Although the publisher is always free to negotiate terms directly with the vendor, in most cases the model contract will be adopted, eliminating the need for many time-consuming bilateral negotiations between the vendor and individual publishers. As a result, the number of participants and the size of the database ramps up much more quickly than it would otherwise.

Current status

During the first six months of the pilot, 3.3 million publications were indexed and participants executed roughly 2,300 document checks. An informal survey of the participants, which included publishing staff as well as journal editors, mid-way through the pilot, revealed that 100% of participants who responded would include their content in a CrossCheck production service, and 91% expected to use such a paid service to check submissions for plagiarism. On the implementation side, the pilot experience has highlighted the importance of integrating CrossCheck into the various manuscript tracking platforms currently in use. Although there is still critical work to be done on CrossCheck's infrastructure, the production service may well launch later this year.

There is certainly no question about the demand for the service among academic publishers, editors and authors, and the prominent launch of a cross-publisher initiative to verify the originality of submitted works will go a long way toward deterring plagiaristic submissions.

Reference

1. Errami, M and Garner, H, *Nature*, 2008, 451, 397–399. | DOI:10.1038/451397a.

Article © Amy Brand

To view the original copy of this key issue, published in *Serials*, the journal of the UKSG, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=21&issue=1&page=54>

For a link to the table of contents for the issue of *Serials* in which this article first appeared, click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=21&issue=1>