Control C, Control V: Plagiarism Detection in a Web 2.0 World

Moderator:
Kirsty Meddings
CrossRef
Oxford, United Kingdom

Speakers:
Bernard Rous
Association for Computing Machinery
New York, New York

Harold "Skip" Garner
University of Texas
Southwestern Medical Center
Dallas, Texas

Harvey Marcovitch
Committee on Publication Ethics
Balscote, United Kingdom

Reporter:
Roma Subramanian
Texas A&M University
College Station, Texas

This session focused on the methodological and technological approaches used by three organizations to identify and handle cases of plagiarism.

Bernard Rous, deputy director of publications at the Association for Computing Machinery (ACM), began the session by describing ACM's plagiarism policy. (ACM publishes 42 journals and magazines and is the oldest and largest computer society in the world.) Rous said that the purpose of the policy, published in 2006, was to deal with cases of plagiarism "systematically" and "fairly" and thus "reduce the risk of lawsuits". The policy defines four types of plagiarism: copying large portions of text, copying elements (for example, equations and illustrations) without citation, citing sources but not using quotation marks to identify sourced material, and self-plagiarism. Rous described the step-by-step process that ACM uses to investigate plagiarism cases, how findings are com-municated, the appeals process, and the penalties administered.

Rous is also a member of the board of directors of CrossRef (an organization founded and directed by publishers to enable permanent links for scholars to use to get from any reference to its source document across publishing sites). He described the phases of development of CrossCheck, CrossRef's plagiarism-detection service. By using a Web-based tool to compare the work of an author with a database of scholarly literature, CrossCheck identifies highly similar documents for further investigation. The literature in the database is contributed by CrossRef publishers.

Rous concluded by stating that it was important to take into account both cultural differences and intentionality in evaluating cases of plagiarism.

Harold Garner described how a tool called eTBLAST (electronic Text Basic Local Alignment and Similarity Tool), developed by his laboratory for searching biomedical literature, is being used to study different types of scientific misconduct (such as plagiarism and falsifying research data). Instead of using keywords, eTBLAST uses a piece of text to identify documents in a selected database that are similar to the text submitted to the search engine. The identified documents are ranked by their degree of similarity to the submitted text. By using MEDLINE citations (which include the titles and abstracts of articles) and eTBLAST, Garner and his team created "Déjà Vu: a database of highly similar and duplicate citations". Garner explained how the database permits repeat plagiarism offenders to be identified.

Garner also described how the database was used to carry out a study in which an e-mail questionnaire was sent to the editors and authors of 206 pairs of original and potentially plagiarized papers with copies of both papers. Papers that had the highest similarity in the Déjà Vu database were selected. The areas of high similarity in the seemingly plagiarized papers were marked. The questions—for example, "Were you aware of the later article?"—hinted at the possibility of plagiarism. Garner shared some surprising findings of the study, one of them being that most authors were unaware that their work had been duplicated.

By presenting cases of plagiarism, Harvey Marcovitch, immediate past-chair of the Committee on Publication Ethics (COPE), involved the audience in discussing how editors should correct and combat plagiarism. He stressed that journals' instructions for authors should clearly describe how to present original and cited work. Also discussed was how to handle plagiarism among junior researchers in developing countries, whose unfamiliarity with English may result in inadvertent copying. Marcovitch concluded by providing handouts of a COPE flowchart that suggests actions that editors may take if they suspect plagiarism. Seventeen COPE flowcharts, each describing steps to be followed for a particular type of suspected publication misconduct, are available at *publicationethics.org/flowcharts*.