

March 2004-04-27

The Impact of Technology on Plagiarism Prevention and Detection:
Research Process Automation, a New Approach for Prevention.

James Douglas Beasley

Uniting Networks Incorporated

200 Ashford Center North

Suite 370

Atlanta, Georgia USA

30338

jbeasley@powerresearcher.com / 00-1-770-390-3411

Copyright © 2004, Uniting Networks Incorporated. All Rights Reserved.

Abstract

For the past several years, technology approaches to plagiarism prevention have focused on detection, meaning the ability for a lecturer to utilise software to analyse student papers and detect whether or not content was plagiarised from another source. This paper proposes adding another approach, Research Process Automation, which focuses on automating elements of the research and writing process and more specifically on the development of research work products. This approach provides plagiarism prevention through just-in-time guidance, research project management, productivity tools, and tracking features. The topics included are the scope of the plagiarism problem, current approaches, some of the strengths and limitations of detection technology, the advantages of using Research Process Automation in plagiarism prevention, the causes of plagiarism, and how Research Process Automation and Research Development Environment software can address various causes of plagiarism.

Scope of the Plagiarism Problem and Current Approaches

In the last several years, plagiarism in higher education and further education, particularly Internet or 'cut-and-paste' plagiarism from digital sources has increased in frequency. In the United States, a 2003 Rutgers University study of 18,000 students, 2,600 faculty, and 650 teaching assistants on 23 campuses found that "thirty-eight percent of the undergraduate students completing the survey indicated they had engaged in one or more instances of cut & paste plagiarism using the Internet in the past year - paraphrasing or copying a few sentences of material from the Internet without citing the source. This is a dramatic increase from the 10% who acknowledged 'cut & paste' plagiarism using the Internet in a similar survey conducted only two years ago" (New Study Confirms Internet Plagiarism Is Prevalent 2003). Historically, the approaches to combating plagiarism have focused on education and honour code programs, and detection of plagiarism after it has occurred. Detection, if done manually, can be a very labour intensive process and may be impossible to conduct within time and personnel resource constraints. The advent of plagiarism detection technology was of great benefit for institutions and lecturers that can effectively utilise the technologies.

Plagiarism Detection Technology—Strengths and Limitations

Detection software products use various methods to accomplish their analyses including text matching with indexed sources, style analysis of content, and tests requiring students to fill in blanks in their own paper. Each of these methods has advantages and disadvantages but all are similar in that they attempt to detect plagiarism after it has been committed and rely on the threat of detection, presumably coupled with punishment, as a deterrent. While this approach certainly equips educators with more time-effective tools

to identify plagiarism and the fear of detection may deter improper behaviour, there are issues with present detection technology.

Detection technology based on matching text from previous student papers and/or the publicly accessible Internet has inherent limitations as noted in the study by Weinstein and Dobkin of the University of California at Berkeley. Their study, which used the TurnItIn.com algorithm, mentions three possible reasons for failure to match information from the Internet that is used in a paper. “First, web sources may have disappeared between when they were cited in the paper and when the paper was checked. Second, since no image of the web is complete and necessarily lags behind the current state of the Internet, it is possible that a web source used in a paper is a new addition to the web or in part of the web not searched by the algorithm. Finally, the algorithm does not have access to all subscription sources of information on the web” (Weinstein, Dobkin 2002:26). In reality, only a small fraction of total Internet content can be analysed. The ‘deep web’ is much larger than the public web. The *How Much Information? 2003* study at the University of California at Berkeley found that “the surface web is about 167 terabytes as of summer 2003; BrightPlanet estimates the deep web to be 400 to 450 times larger, thus between 66,800 and 91,850 terabytes” (How Much Information? 2003).

Even plagiarism detection services that also scan some subscription databases are not comprehensive. Lecturers and department heads in the U.S. tell us of entire departments or institutions that do not use text-matching detection technology specifically because of its inability to scan for matches in subscription databases such as Lexis-Nexis, EBSCO, Swets, and others. The Weinstein and Dobkin study elaborated on the effects of the limitations noted previously and estimated a TurnItIn.com “detection percentage of web sources... [of] 63%” which indicates that “all measured plagiarism rates should be multiplied by a factor of $1/0.63=1.6$ ” (Weinstein and Dobkin 2002:27).

Other factors that may inhibit the use or effectiveness of plagiarism detection technologies might include concern over the legal implications of submitting student work to a third-party vendor and sensitivity to a possible ‘guilty until proven innocent’ feeling between students and lecturers. The latter issue is perhaps more prevalent with the use of the Glatt Plagiarism Screening Program software which requires students to prove that their work is original by supplying missing words in their paper during a testing session.

In addition, text-matching technology may not help differentiate the inadvertent or accidental plagiarist from the intentional plagiarist as it provides the lecturer with few, if any, insights as to the intent of the offender.

Deterrence

Even with text-matching analysis limitations, there is evidence that text-matching detection technology provides a deterrent effect. The Weinstein and Dobkin study found that their control classes—those that were not told their papers would be checked by plagiarism detection software—had “an average

plagiarism rate of 17.2%” and the treatment classes—those who were warned that their papers would be checked—had “an average plagiarism rate of 10.7%”. When the researchers used their “internal calibration to provide an estimate of the true, underlying plagiarism rate” they found that “approximately $17.2\% \times (1/0.61) = 28\%$ of student papers are plagiarized” (Weinstein and Dobkin 2002:36). Applying this same calibration to the treatment classes would yield a plagiarism rate of approximately 17.5%. Although the deterrent effect is significant, the data shows that the awareness of scrutiny does not always work as a deterrent and there is opportunity for improvement.

Research Process Automation

Given the limitations of detection technology and identified barriers to use, an additional technological approach can fill the gaps. Research Process Automation is an approach to plagiarism reduction that focuses on prevention. We use this phrase to describe a discipline that seeks to automate multiple parts of the research and writing process, which we describe as “education, research, writing, citation, submission, review, and publication” (Beasley 2003). In its most holistic sense, Research Process Automation can encompass technologies such as courseware, search engines, databases, summarization tools, writing applications, bibliographic management tools, and plagiarism detection software. Within the Research Process Automation framework, we use the term Research Development Environment, or RDE, to describe integrated, multi-function software that enables or facilitates the creation of research work products. This can include functionality that supports primary as well as secondary research. Thus, Research Process Automation describes an approach or discipline while RDE describes a specific software implementation within that area.

Research Development Environment Software

Until recently, little was done to provide technology tools to researchers and students that guided them in correctly conducting research and increasing their productivity while doing so. RDE software is that which enables, organises, accelerates, and monitors the research and writing process, and creates a better communication vehicle between colleagues or lecturer and students. In the case of a lecturer and students, the work products passed between them can be the finished paper and all digital work artefacts relevant to the assignment. This might include research project plans, concept maps, offline copies of web pages, notes, annotations, source information, research and writing activity logs, and presentation materials. Jeffrey D. Reynolds, Director of Product Development at Uniting Networks Incorporated, and I formulated the idea for RDE software in response to what we saw as problems and inefficiencies in Internet research and writing tasks. As the idea for an RDE grew, it became apparent that principles from other disciplines could be applied to the challenge. Project Management, Business Process Management, and Software Engineering are three key disciplines that contribute to the RDE concept. The RDE tool produced by Uniting Networks Incorporated is called PowerResearcher™.

Principles from Project Management, Business Process Management, and Software Engineering

Project Management principles can be applied to the research and writing process to achieve more desirable outcomes. As discussed later, disorganization is one of the root causes of plagiarism. Lecturers we have spoken to mention this as one of the most common causes. Two central project management principles are project planning and monitoring. Kerzner defines the elements of project planning as “definition of work requirements, definition of quantity and quality of work, and definition of resources needed” while project monitoring includes “tracking progress, comparing actual outcome to predicted outcome, analyzing impact, and making adjustments” (Kerzner 1995: 3). Many of these elements are relevant in the context of a research and writing project or assignment. In order to meet deadlines and quality standards, students should monitor their progress and continually reassess the time and effort needed to complete the assigned work. This becomes more important when team-based assignments are undertaken involving multiple people who contribute to the final project. Of course, the planning and monitoring processes and tools that are appropriate to a large project with many people and interdependent tasks are unwieldy for a project involving a single writer. However, the basic principles of planning and monitoring are still applicable. Research Process Automation includes planning and monitoring functions specifically designed for research and writing tasks for both a team and individual.

Organisations use Business Process Management techniques and tools to improve efficiency and quality. These methods can also improve the research and writing process. Planning and monitoring gives students the ability to evaluate the time they spend on a particular task, the ratio of time spent on different tasks, and the sequence of tasks. Students can then compare their plan to the actual effort, time, and sequence of tasks during the project and after the project is complete. With this information, students can identify areas for improvement and thus refine their research and writing methods. The lecturer can also review monitored activities and instruct students on better techniques or identify students that need remediation. The lecturer may also gain insights that result in improvements in their teaching methods.

Interestingly, there are parallels between utilizing software productivity tools for the research and writing process and for Software Development. In the early days of computer programming, programmers used simple text editors to write software code by hand. In time, vendors created software-editing tools that were more sophisticated and which helped programmers to format proper syntax. Eventually, Integrated Development Environment, hereafter referred to as IDE, tools were created that combined several functions necessary for efficient software development into a common user interface such as an editor, compiler (for converting code into an executable format that can run on a computer), and a debugger. The advantages of an IDE are increased productivity, quality, and traceability. In the software context, traceability is the capacity to track logical connections from business and functional requirements through programming elements to finished software. Similarly, a Research Development Environment tool can provide similar advantages to the researcher and writer. The advent of digital forms of research,

particularly the Internet, make many of the tasks and tools involved in research more akin to software development than in the past. It makes sense, therefore, to use lessons learned from the software industry to enhance research and writing.

Proactive Plagiarism Prevention

Of course, the most important aspects of Research Process Automation and an RDE tool to this discussion are plagiarism prevention. This can be accomplished in four ways:

1. The student or researcher finds that productivity gains allow them the time to follow research best practises. Productivity gains can come from several functions including project planning and monitoring, source logging, annotation, and citation creation.
2. The student or researcher is less likely to make a reference mistake due to automatic logging and tracking of sources and formatting of citations.
3. The ability of lecturers to review time-based logs of research and writing activity lets them gain insights into students' research and writing activities, work habits, and time expended to verify that the students' output matches the recorded effort.
4. The increased ease and efficiency of research and writing changes the 'risk versus reward' and/or 'cost versus benefit' calculations in the minds of students or researchers that contemplate committing plagiarism.

Lecturer and Student Communication—a 360 Degree Assignment Model

The phrase '360 Degree Assignment Model' describes a process for using a Research Process Automation/RDE tool by both lecturer and student. In the 360 Degree Assignment Model, the lecturer uses a Research Process Automation tool to prepare an assignment starter project with teaching materials that are passed in digital form to their students. These materials could include lesson plans, assignment instructions, sample web pages, templates, citation examples, an assignment schedule or plan, and research evaluation criteria. A variety of digital content can be packaged for easy distribution to students. Students can then use the lecturer's assignment starter project as a starting point for their own work. By adding their work to the project, students can complete their assignment work and always have their instructor's instructions and examples at their fingertips. On completing their work, students hand in their project file digitally to the instructor via e-mail, removable media, network, or web upload. The lecturer then has every piece of information pertaining to the assignment at hand. This gives the lecturer access to more information about the student, and their intentions, than ever before and can save many hours of reference checking since collected digital research resides with the student's work.

Causes of Plagiarism and Mitigation

In order to understand how Research Process Automation and RDE software can help reduce plagiarism; it is helpful to discuss some of the causes and motivating factors of plagiarism and the specific RDE tool features that can reduce plagiarism. Many of the causes of plagiarism are not due to ethical misconduct. The Weinstein and Dobkin study states that half of the control plagiarism rate observed in the study was the “result of inattention by the students” (Weinstein and Dobkin 2002).

1. Disorganization

This may include poor note taking, inadequate time management, underestimation of workload, or the inability to choose from many sources. Research Process Automation/RDE tools can provide such features as integrated research project planning, task and deadline reminders, progress monitoring, automated source tracking, information capture and storage, and concept/topic-based organisation.

2. Information Overload

The student may feel they have too many sources to evaluate or they may be unable to choose a set of information, which results in spending too much time collecting, rather than analyzing, information. The Internet contributes to this issue due to the vast diversity of information and the relative ease with which one can find it. Research Process Automation provides frameworks that facilitate source search, collection, and organization. By speeding this process, more time can be devoted to information analysis.

3. Ethical Lapses

Attitudes about the ethics of plagiarism and, indeed, all forms of cheating have degenerated steadily. The 29th *Annual Survey of High Achievers by Who's Who Among American High School Students* found that 80% admitted to cheating on school work—up 4% from the previous year—and “53% said the transgression was no big deal” (Who's Who 1998). Although Research Process Automation cannot, by itself, inculcate a sense of ethics into students in whom it is lacking, the guidance and productivity gains offered may change the ‘risk versus reward’ equation in their minds and encourage them, out of simple self-interest, to properly conduct research.

4. Laziness

This includes a simple unwillingness to work diligently, procrastination, and the inclination to do the least amount of work necessary. If laziness is a motivating factor, the current climate of easy access to information that can be stolen provides ample temptation. A Research Process

Automation/RDE tool, by saving the student time, may mitigate the natural inclination of some individuals to avoid work.

5. Ignorance

Students can be uninformed, or misinformed, about the definition of plagiarism, when to cite, what to cite, and how to cite. Research Process Automation/RDE tools can include research and citation features that instruct as well as automate. Reminding the student of correct procedures at the exact moment of research activity can increase the retention of learned procedures and alleviate the anxiety associated with uncertainty.

6. Fear

Students may cheat out due to fears of inadequacy, failure, and self or parental expectations. There may also be a defensive element. Harris observes, "Cheating in self-defence may appear rational in a highly competitive atmosphere, especially where students believe there are few operative punishments" (2001:6). Research Process Automation/RDE tools may help to inspire confidence in students as to their abilities to complete a quality assignment. While these tools cannot, and should not, replace the confidence created by knowledge, some students may benefit from a tool that automates some of the manual and more repetitive tasks.

7. Cryptomnesia

The phenomenon of 'forgotten knowledge' may occur when the researcher fails to keep careful track of information gathered and utilised. Harris notes that students that "engage in heavy research (as, say, graduate students do) ... may be setting themselves up for cryptomnesic plagiarism unless they very carefully record and track every source" (Harris 2001:12). The sheer amount of information available today may be a key contributor to this problem. The *How Much Information? 2003* study at the University of California at Berkeley found that, in 2002, the world produced five exabytes of new information in print and on film, magnetic, and optical storage media. The report noted that "five exabytes of information is equivalent in size to the information contained in half a million new libraries the size of the [U.S.] Library of Congress print collections" (How Much Information? 2003). Of course, one of the central tenets of the Research Process Automation concept is automatic tracking of sources. This alone, if used consistently, can reduce the possibility of cryptomnesia.

8. Thrill Seeking

Some students may even relish the excitement of breaking rules and avoiding censure. While a certain percentage of students have probably fallen into this category in the past, the opportunities available today are greater than ever. While Research Process Automation cannot

alter this personality trait, the ability for the lecturer to audit research activities and match those to the resulting student work makes this a time-consuming, costly, and risk-filled endeavour.

Types of Plagiarists

After considering some of the causes of plagiarism, a picture of three broad types of plagiarists emerge.

1. The accidental plagiarist is one who either does not understand plagiarism or makes a mistake in quoting, citing, or paraphrasing (Beasley 2003). The accidental plagiarist may suffer from ignorance, information overload, disorganization, and/or cryptomnesia.
2. The opportunistic plagiarist is one who knows that it is wrong to plagiarise but does it anyway due to disorganization, information overload, ethical lapses, laziness, and/or fear.
3. The committed plagiarist is one who intends, with forethought, to cheat by stealing others words and/or ideas. This is the type of plagiarist who might purchase a paper from a paper mill. Committed plagiarists may suffer from ethical lapses, fear, or even thrill seeking. The other factors mentioned above can also contribute to this behaviour (Beasley 2003).

As was noted previously, detection technology cannot fully address the issue of accidental plagiarism because it cannot determine whether the plagiarism is truly accidental or an effort to disguise cheating. The fear of detection and punishment may deter the opportunistic plagiarist, but this is not certain. For both accidental and opportunistic plagiarists, Research Process Automation can help mitigate the issues faced by these groups and, thus, reduce plagiarism. If the lecturer utilised the 360° Assignment Model described previously, even committed plagiarists would face the time-consuming and technically difficult task of faking their research and writing activities.

Issues for Further Study

The Research Process Automation concept is relatively new and is currently being utilised in some higher education institutions and secondary schools in the United States. A treatment/control study would help to test and quantify the potential benefits and to refine concepts and design parameters. Data migration and/or functional integration points between existing technologies and products can also be explored to maximize benefits and cost effectiveness.

Conclusion

No single process or technology can address the numerous varieties of plagiarism and the corresponding motivations for committing the offence. It is our belief that the addition of Research Process Automation and Research Development Environment software can prevent or reduce plagiarism or fill gaps in existing approaches by providing just-in-time guidance, research project management, increased productivity, and information tracking features. Overall, it is our opinion that a combination of education and honour code programs, Research Process Automation RDE software, and plagiarism detection technology is the best strategy to combat plagiarism in the Internet age.

References

- *About the Project Management Profession*. URL: <http://www.pmi.org/> [3 April 2004]
- *Cheating and succeeding: record numbers of top high school students take ethical shortcuts: Who's Who Finds Troubling Trends, Some Good News*. (1998) URL: <http://www.whoswho-teachers.com/3attitudeANDsopinions/29.aspx> [29 February 2004]
- *How Much Information? 2003*. (2003) URL: <http://www.sims.berkeley.edu/research/projects/how-much-info-2003/> [29 February 2004]
- *New Study Confirms Internet Plagiarism Is Prevalent*. (2003) URL: <http://ur.rutgers.edu/medrel/viewArticle.html?ArticleID=3408> [8 October 2003]
- Beasley, J.D. *Research Process Automation Summary*. (2003) URL: http://www.powerresearcher.com/rpa_summary.html [28 February 2004]
- Harris, R. (2001) *The Plagiarism Handbook*. Pyrczak Publishing.
- Kerzner, H. (1995) *Project Management, A Systems Approach to Planning, Scheduling, and Controlling*. Fifth Edition. Van Nostrand Reinhold.
- Weinstein, J. and Dobkin, C. (2002). *Plagiarism in U.S. Higher Education: Estimating Internet Plagiarism Rates and Testing a Means of Deterrence*. URL: <http://webdisk.berkeley.edu/~Weinstein/Weinstein-JobMarketPaper.PDF> [12 April 2004]