Plagiarism and the Web: Myths and Realities

An Analytical Study on Where Students Find Unoriginal Content on the Internet

turnitin®

Prevent Plagiarism.
Engage Students.

Table of Contents

1.0 Summary

The move from an analog to a digital culture is raising a new set of challenges for educators and students when it comes to writing and critical thinking. The following study examines what web sources students rely on and how educators can help students develop better citation and writing skills.

Key findings include:

- **Plagiarism is going social**
  One-third of all content matched in the study is from social networks, content sharing or question-and-answer sites where users contribute and share content.

- **Legitimate educational sites are more popular than cheat sites**
  One-quarter of all matched material is from legitimate educational web sites, almost double the number that comes from paper mills or cheat sites.

- **15 percent of content matches come directly from sites that promote and benefit from academic dishonesty**
  Paper mills and cheat sites are the third most popular category for matched content.

- **Wikipedia is the most popular site for matched content**
  Wikipedia remains the most popular single source for student-matched content on the Web, comprising seven percent of matches in the months examined.

- **Educators with the knowledge and tools can address the growing problem**
  Educators who employ the proper tools and technologies can significantly mitigate plagiarism. For example, institutions with widespread adoption of Turnitin see a reduction in unoriginal content of 30 to 35 percent in the first year. By the fourth year, many institutions see levels of unoriginality fall by up to 70 percent.

2.0 Introduction

In the last 15 years, the rise of the Internet has changed the way students discover, use and source content for writing. Students now have more information available to them than was previously imaginable. While this wealth of information brings benefits in terms of the dispersion and accessibility of educational content, it can also lead to poor or dishonest writing practices that impede the learning process. In short, if used incorrectly, digital content can hamper the ability of students to write and think critically.

The following study examines what web sources students rely on for unoriginal content in their written work. It is based on close to 40 million student papers that were submitted to Turnitin, the leading plagiarism prevention and writing application for instructors and students. The study examined and classified 140 million content matches in the student papers to provide an unparalleled glimpse into the resources students use for content on the Web.

The Turnitin Database

Papers submitted to Turnitin are compared against three main repositories of information: 14 billion current and archived web pages; 150 million student papers; and 110 million content items from academic publishers.

The study reveals some surprising results about student behavior when it comes to source material. Contrary to general perceptions, the vast majority of students who have matched content in their work do not rely on cheat sites or paper mills. Instead, many more are using legitimate homework, academic and educational sites as research sources. The study also shows that student research and writing practices are following similar trends of the Internet as a whole. Increasingly, students rely on social networks and user-generated content sites such as content sharing and question-and-answer sites to find materials that they include in their papers.

The report outlines some broad trends based on the findings of the study and offers instructors, administrators and parents steps to take to help students use and document sources from the Web.

3.0 Popular Content Resources on the Web

Turnitin classified the top 100 most popular web sites that matched existing content in the Turnitin databases over a period of ten months (June 2010 to March 2011).

The categories are as follows:

Social Networking and Content Sharing
This category encompasses sites that rely on user-generated content rather than professionally published content. The sites include social networking sites such as Facebook and MySpace, content sharing sites such as Scribd and SlideShare and Question & Answer sites such as Yahoo Answers and Answers.com.

Homework and Academic
The next category consists of academic, educational and homework help sites that offer a range of content to educate students. Many of these sites carry .org or .gov domain registrations such as medlibrary.org and the National Institute of Health web site www.nih.gov. It also includes web sites dedicated to helping students prepare for tests or aid in homework assignments, such as www.coursehero.com and www.bookrags.com.

News and Portals
Unlike the social and user-generated content web sites, this category consists of sites that employ a traditional publisher model. Examples include The New York Times, The Huffington Post and Associated Content.

Paper Mills and Cheat Sites
There are a large number of sites that are geared toward cheating and academic dishonesty. Whether under the guise of paper sharing or paper sales, these sites seek to profit from student need by either selling, exchanging or placing advertising around offers of original student papers. Sites include www.oppapers.com and www.allfreepapers.com. What many students fail to appreciate is that these papers are either a) indexed by the Turnitin web crawler or b) have already been submitted to the Turnitin database in someone else's name.

Encyclopedias
Encyclopedia web sites were classified into their own category because of the tremendous popularity of Wikipedia as a student source on the Web. Other encyclopedia sources include Brittanica.com and Encyclopedia.com.

Other
Finally, there were a few sites whose origins or purpose could not be determined because the site is no longer operable. These sites were classified into an "Other" category for purposes of this analysis.

Matched Content vs. Plagiarized Content

There is a lot of confusion about what constitutes plagiarized content. This study examines content that matched items previously stored in the Turnitin database.

Turnitin determines if text in a paper matches text in any of the Turnitin databases. The service does not detect nor determine plagiarism – it detects patterns of matching text to help instructors determine if plagiarism has occurred.
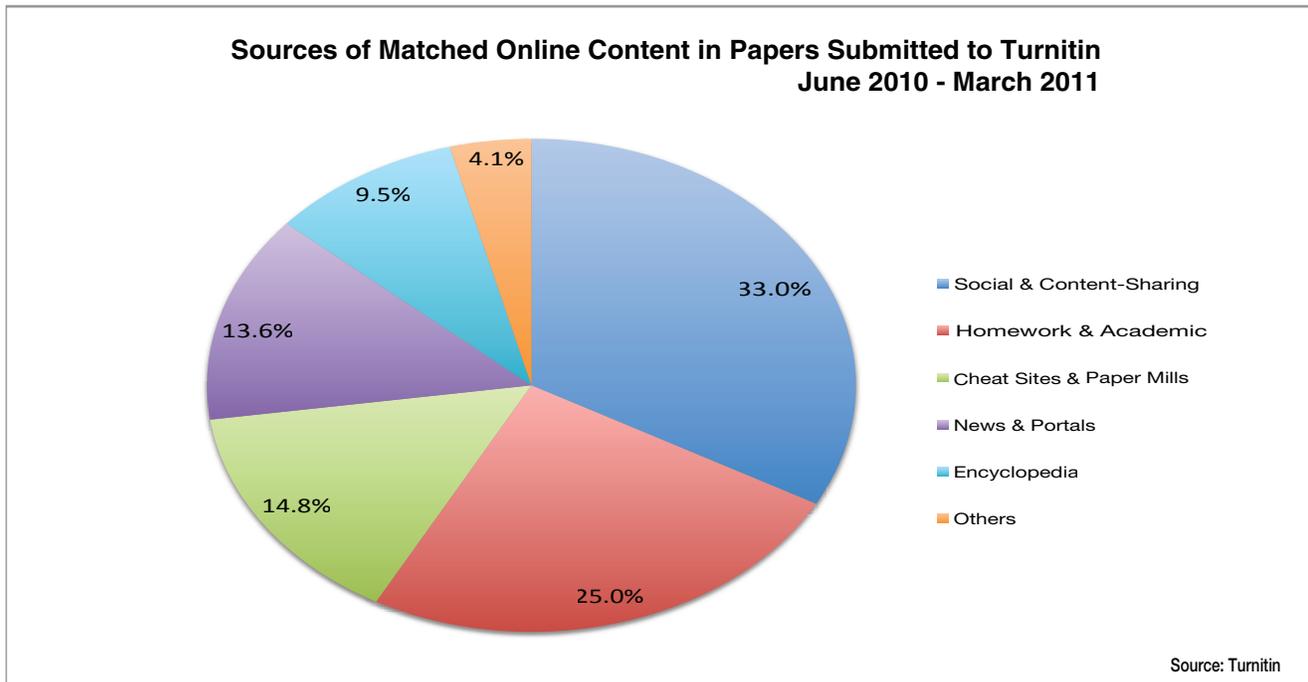
Indeed, the text in the student's paper that is found to match a source may be properly cited and attributed. It is recommended that instructors carefully review the Originality Report and all matches before making any determination of plagiarism. Such determinations of plagiarism require human judgment, and instructors and students alike should understand their institution's academic integrity policies before turning in written assignments.

Source: Turnitin FAQs
http://turnitin.com/static/products/faqs.php

4.0 The Most Popular Student Sources

SO WHAT WERE THE MOST POPULAR STUDENT SOURCES BY CATEGORY?

**Sources of Matched Online Content in Papers Submitted to Turnitin
June 2010 - March 2011**

- Social & Content-Sharing
- Homework & Academic
- Cheat Sites & Paper Mills
- News & Portals
- Encyclopedia
- Others

33.0%
25.0%
14.8%
13.6%
9.5%
4.1%

Source: Turnitin

Social and content sharing web sites comprised the highest percentage of all matched content over the course of the study. Legitimate homework and academic help sites were second, followed by cheat sites/paper mills and news and portals a close third and fourth. The fifth most popular category was encyclopedia.

5.0   The Top Eight Most Popular Sites for Matched Content

Wikipedia remains the most popular single source for student-matched content on the Web, comprising seven percent of matches in the months examined.

The top eight matched sites, along with their given category, are:

1) en.wikipedia.org - Encyclopedia

2) answers.yahoo.com – News & Portal

3) www.answers.com - Social & Content Sharing

4) www.slideshare.net - Social & Content Sharing

5) www.oppapers.com - Cheat Sites & Paper Mills

6) www.scribd.com - Social & Content Sharing

7) www.coursehero.com - Homework & Academic

8) www.medlibrary.org - Homework & Academic

In addition to Wikipedia's popularity, it is interesting to note that only one of the top eight most popular sites is dedicated to helping students cheat by providing unoriginal content. Indeed, the study shows that out of the 25 most popular sites, 14 are legitimate student resources for learning.

Should Students Cite Wikipedia?

While there is no doubt that Wikipedia is a useful resource for almost any topic imaginable, many instructors discourage, even reject, papers that rely on Wikipedia as a source because they question the accuracy and validity of the content. Instead, instructors are asking students to cite primary or secondary source material.

6.0   Key Findings

An analysis of the results demonstrates a few simple truths about the state of plagiarism and the Web:

Sharing and Originality are Colliding
A digital culture that promotes sharing, openness and re-use is colliding with one of the fundamental tenets of education – the ability to develop, organize and express original thoughts. For many students who have grown up sharing music, retweeting thoughts and downloading free software, the principle of originality in research and writing can seem antiquated. It is important for educators to draw a clear line between what can be reproduced and what must be created. If not, there is a risk that a generation of students will not develop the critical thinking and communication skills necessary for a productive life.

Most Students are Trying to do the Right Thing
While the study makes clear that close to 15 percent of matches come from cheat sites and paper mills, the more encouraging finding is that many more students are frequenting legitimate academic or educational web sites. It is also worth noting that in the ten-month period, sites promoting academic dishonesty saw the largest decline of any category. While it is too early to draw long term conclusions, the short-term results are encouraging.

Understanding Originality

For each paper submitted, Turnitin provides educators with a color-coded version of the paper that clearly highlights any content that matches other sources. Educators can then review those areas to determine if the content is cited correctly and if not, can go back to the student and take appropriate action. It is rare that papers are 100 percent unoriginal, but when that is found to be the case, Turnitin provides the evidence instructors need to address the issue.

7.0 Recommendations for Educators

There are three simple steps educators can take to ensure students use and cite the right sources on the Web.

1. **Teach Proper Citation**
   In order to better teach sourcing, educators can instruct students on how to take what they have learned from their sources and either summarize, paraphrase or quote it. Furthermore, educators can guide students in proper integration and citation practices. Parents can have an active role in promoting original work by helping teach the art of careful note taking and instilling the value of academic integrity in their children.

2. **Use Technology to Address a Digital Problem**
   Educators can teach students digital literacy by investing in tools and technologies that flag potential plagiarism. Relying on instructors to "Google" phrase after phrase of student writing is not only taxing on already overburdened teachers, but it will not provide an overall context of how much of the student paper is unoriginal nor will it offer an easy way to convey the findings. Turnitin is designed to assist educators in not only identifying potential issues of plagiarism but also helping instructors provide students with rich feedback during the writing process.

3. **Create Teachable Moments**
   Educators can teach plagiarism before it occurs. Demonstrating educational tools, such as Turnitin and plagiarism. org to students at the beginning of the semester or school year can deter students from improper research, citation and writing practices. Even more, showing students a detailed Originality Report of their written work is a powerful instructional tool that creates a "teachable moment" where the student's work is at the center of the writing and assessment process.

Do Plagiarism Prevention Services Work?

A recent in-depth statistical analysis of more than 21 million submissions to Turnitin uncovered what 10,000 customers already know – Turnitin improves the level of originality in student writing. Institutions with widespread adoption of Turnitin see a reduction in unoriginal content of 30 to 35 percent in the first year. By the fourth year, many institutions see levels of unoriginality fall by up to 70 percent.

8.0 Conclusion

This report outlined some of the major trends regarding how students find and source material from the Internet. The move from an analog to a digital culture demands a new approach from educators. Instructors, administrators and parents need to combine basic pedagogy with the tools and technologies to educate students about finding and citing the correct sources in their written work.

About Turnitin and iParadigms, LLC

iParadigms, LLC is the world's leading provider of web-based solutions for plagiarism prevention. The company's products include Turnitin, used by educators worldwide to check students' papers for originality, to enable web-based peer review and for digital grading of student work. iParadigms' iThenticate solution is the leader in plagiarism detection in research, publishing and many other commercial markets. The company's solutions check millions of documents each month and are used in over 100 countries. iParadigms is headquartered in Oakland, CA with an international office located in Newcastle, United Kingdom. iParadigms is backed by Warburg Pincus. For more information, please visit www.turnitin.com.