



# The Broadest Search

## Turnitin Outperforms Search Engines

### WHY TURNITIN?

Turnitin's plagiarism prevention system is a significant improvement to the shot-in-the-dark approach of using online search engines like Google® to detect student plagiarism. Here's why:

- Our database contains millions of published works unavailable through any online search engine.
- We use proprietary technology to find less obvious instances of plagiarism where the student has used multiple sources or changed parts of the text.
- The speed of our system—it processes hundreds of papers per minute—is in stark contrast to the hours of time it takes instructors to copy search phrases from each paper into Google and evaluate the results.
- The Turnitin Internet database is vast, indexing 4.5 billion pages and updating at a rate of 40 million pages per day. The database includes millions of pages that are no longer available online or on other search engines.
- Our web crawler archives any publicly accessible documents up to 2 MB in size. Most search engines archive documents up to only 100 KB, meaning that Turnitin searches potentially important documents that others ignore.

### COMPREHENSIVE DATABASES

Plagiarized content comes from three main sources: the Internet, published works, and other student papers.

Search engines do not have access to either published works or student papers, so no amount of Googling will help in these cases. And while instructors may have access to a small pool of previously submitted papers and published works, they generally lack the means to perform effective searches on this material.

Turnitin solves these problems with sophisticated search algorithms that compare every paper submitted against three distinct digital databases: a current and extensively archived copy of the Internet that we update daily, a huge database of books and journals, and a database of digital fingerprints of the millions of papers already submitted to us. Turnitin is the only technology that searches against comprehensive databases of the three main sources of plagiarized content.

### BETTER SEARCHING

At peak usage, Turnitin analyzes more than 20,000 student papers every day, and about 30 percent of those papers contain a significant amount of unoriginal material. A major shortcoming of using Google to detect plagiarism is that the vast majority of unoriginal papers are culled from many different sources. In fact, our data indicates that

less than one percent of plagiarized papers originate from a single source.

Detecting multiple-source plagiarism using a search engine can be exceedingly difficult. It generally entails cutting-and-pasting multiple phrases from suspect papers into Google and then sorting out the results—an ineffectual approach compared to the efficiency and breadth of Turnitin's powerful search. Turnitin can even identify unoriginal passages in papers where many of the words have been changed to outwit traditional search engines.

### FASTER SEARCHING

Perhaps most importantly, Turnitin saves valuable instructor time. The opportunity cost to faculty using search engines to police their student work is

*The service allowed me to find the exact paper a student had turned in as his own work. I knew that the paper was plagiarized, but would never have been able to find the source without Turnitin.com.*

—From our most recent user survey

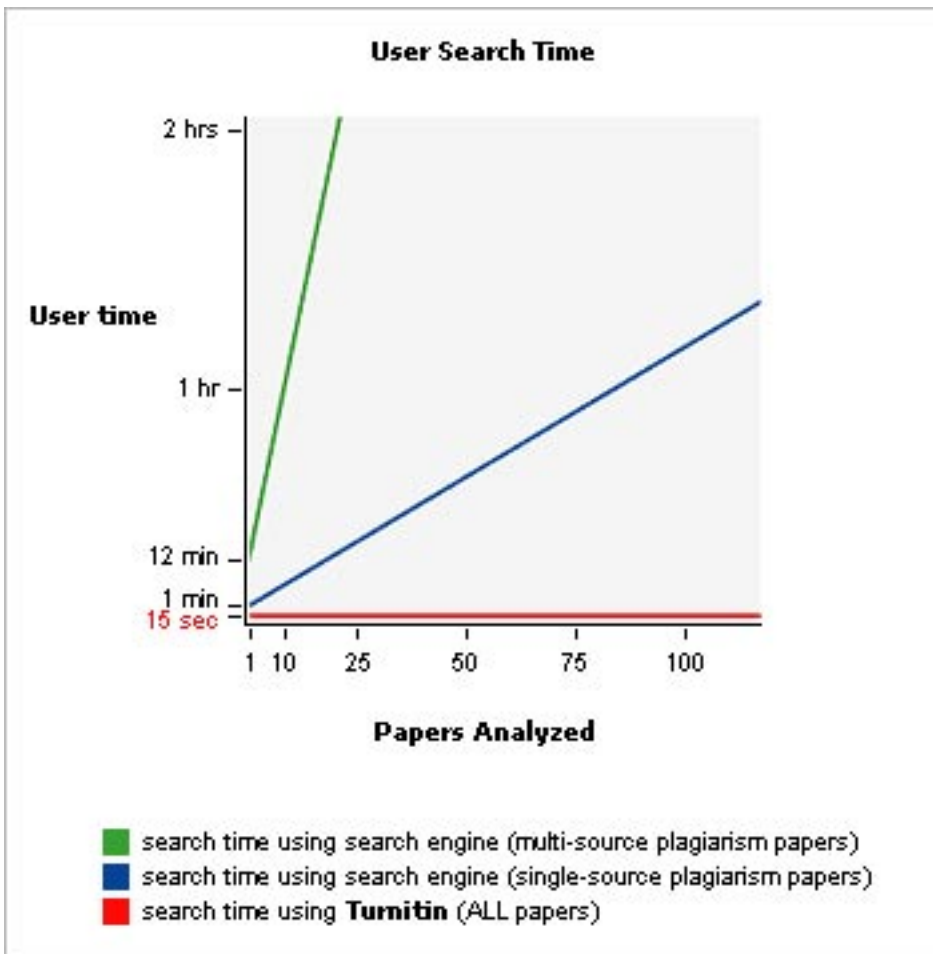
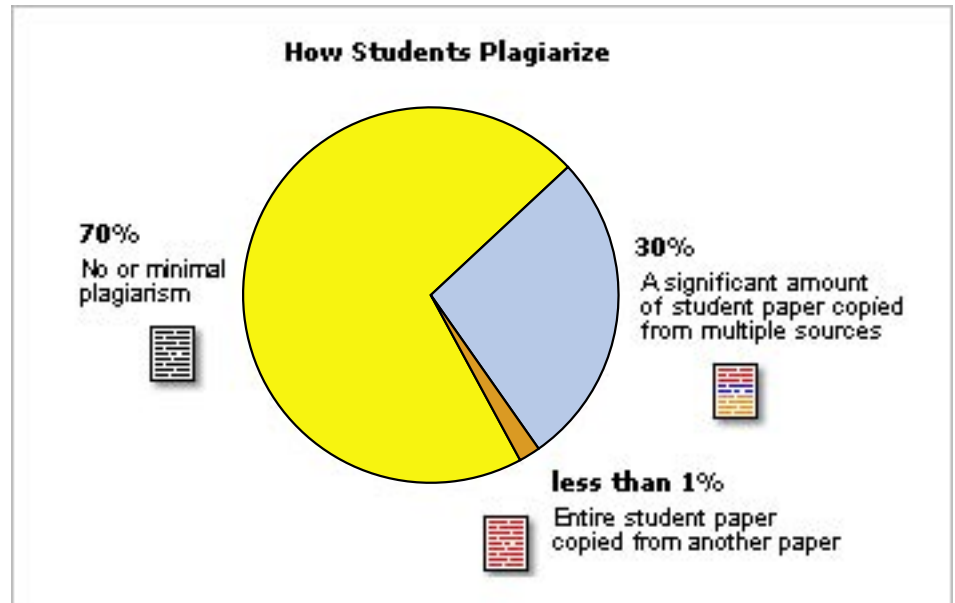
enormous. Searching for the source of one phrase in each of seventy papers can take well over two hours; searching for multiple phrases from each paper quickly turns a cut-and-paste session with Google into a frustrating and unmanageable task.

## QUANTIFIED RESULTS

The figures below are based on data culled from papers submitted to Turnitin, as well as independent studies conducted to determine the average amount of time invested when using search engines to look for plagiarism.

**Fig. 1: How Students Plagiarize**

Based on data culled from papers submitted to Turnitin, this graph indicates that: **1)** a troubling 30 percent of student papers submitted contain a significant amount of unoriginal material; **2)** plagiarized material typically comes from many different sources (blue shading); and **3)** less than 1 percent of the millions of papers Turnitin has analyzed have been completely copied from another paper on the Internet (orange shading).



**Fig. 2: User Search Time**

This graph shows the typical amount of time it takes to search for plagiarism. The green trace shows that papers taken from multiple sources take about 12 minutes each to search using Google. The blue trace shows how long it takes to locate papers that have been copied from a single source. The red trace shows that it takes only 15 seconds to search 125 papers with Turnitin, regardless of their origin.



Since 1996, Turnitin has been helping millions of faculty and students in over fifty countries improve writing and research skills, encourage collaborative online learning, ensure originality of student work, and save instructor time.

iParadigms, LLC  
1624 Franklin St., 7th Floor  
Oakland, California 94612  
tel (510) 287-9720  
fax (510) 444-1952  
www.turnitin.com