

# Student Plagiarism and You: Prevalence, Implications and Solutions

Melissa Holmberg, Electronic Resources Librarian  
Mark McCullough, Reference Coordinator

Memorial Library  
Minnesota State University, Mankato, U.S.A.



# Presentation Objectives

1. Describe 2 studies that tested search engines as tools for detecting plagiarism in electronic graduate theses
2. Discuss the results of both studies
3. Within the context of the 2 studies explore these questions:

How prevalent is the problem of plagiarism?

What are the implications (for faculty, students, institutions, libraries)?

What role can/should librarians play in combating plagiarism?

# Plagiarism Studies-- Background

Why did we pursue the studies?

- Plagiarism in the news
- Examples of plagiarism in theses in our library
- Campus was purchasing Turn-It-In
- Concerns about privacy issues with commercial software
- Preferred more proactive, instructional solution
- Search engines as an alternative?

# Search Engines as Anti-Plagiarism Tools

Study 1: Using the Google Search Engine to Detect Word-for-Word Plagiarism in Master's Theses: A Preliminary Study

Study 2: Anti-Plagiarism Tools: Scirus v. Google

# Study #1

## Questions:

Is Google an efficient and effective tool for detecting plagiarism in Master's Theses?

- Could Google retrieve the type of content typically used by graduate students in theses?
- Could Google be used by theses advisors to detect possible plagiarism quickly – how much time would it require?

# Study #1

## Method:

- 210 electronic master's theses from 2003
- Phrases (7 consecutive words) selected from theses and searched in Google for up to 10 minutes
- Looked for undocumented or improperly documented (referenced but no quotations)
- When matches were found, matched web pages were printed

# Study #1: Results

- ◆ Matches (potential occurrences of plagiarism) found in 27% of theses examined
- ◆ Average time to find a match < 4 minutes
- ◆ Computer Science/Computer Engineering Theses had most matches

# Study #1

Subject	Found	Total	%
Agricultural, Life, & Medical Sciences	6	31	19.35%
Architecture, Construction, Interior Design	2	8	25.00%
Arts & Humanities	1	23	4.35%
Business & Social Sciences	5	27	18.52%
Computer Science/Engineering	17	39	43.59%
Education	2	10	20.00%
Mechanical Engineering & Aerospace	8	21	38.10%
Other Engineering	10	25	40.00%
Chemistry & Physical Sciences	6	26	23.08%

# Study #1

## Extent of matched phrases

Long phrases (7+ words)	23
Multiple long phrases	11
Entire sentence	6
Multiple sentences	7
Entire paragraph	6
Multiple paragraphs	3
Entire publication	1
	57

# Study #1

Institution	Found	Total	%
L	0	1	0.00%
N	0	1	0.00%
V	0	1	0.00%
W	0	1	0.00%
I	0	3	0.00%
E	0	5	0.00%
U	0	5	0.00%
J	0	10	0.00%
K	1	3	33.33%
O	1	9	11.11%
R	1	10	10.00%
D	2	4	50.00%
G	2	8	25.00%
F	3	5	60.00%
C	3	10	30.00%
B	4	14	28.57%
P	5	12	41.67%
Q	5	13	38.46%
A	6	17	35.29%
T	6	21	28.57%
M	6	33	18.18%
S	12	24	50.00%
	57	210	27.14%

Found matches among 63.63% of institutions

# Study #1

## Conclusions?

- Promise as a tool in terms of time and effectiveness
- Difficult to determine plagiarism with certainty

# Study # 2

## ◆ Questions:

- Same as Study #1 (Further exploration of efficiency and effectiveness of search engines for detecting plagiarism)
- How does Google compare to Scirus?

# Study #2

## Method:

- Sample of 2003 electronic science & engineering master's theses from Worldcat
- 68 theses were checked by the authors
- Phrases (7 consecutive words) selected from theses and searched in Google for up to 10 minutes
- All theses searched in Google and Scirus
- Authors searched independently
- Matches cross-searched using both search engines

# Study #2

- ◆ Scirus searcher found about as many matches as Google searcher
- ◆ Scirus searcher matched against more journal articles
- ◆ Both Search engines found unique matches

# Average Time in Minutes

<b>Scirus</b>	<b>3.79</b>
<b>Google</b>	<b>4.88</b>
<b>Average of Both</b>	<b>4.42</b>

# Types of documents found

Scirus			Google			Total	
Corporate Website	0 of 32	0%	Corporate Website	4 of 34	11.76%	Corporate Website	4
Dissertation	1 of 32	3%	Dissertation	2 of 34	5.88%	Dissertation	3
Government Documents	4 of 32	12.50%	Government Documents	3 of 34	8.82%	Government Documents	7
Journal Article	12 of 32	37.50%	Journal Article	3 of 34	8.82%	Journal Article	15
Magazine Article	0 of 32	0.00%	Magazine Article	1 of 34	2.94%	Magazine Article	1
Patent	1 of 32	3%	Patent	0 of 34	0%	Patent	1
Preprint Server	0 of 32	0%	Preprint Server	1 of 34	2.94%	Preprint Server	1
Proceedings	0 of 32	0%	Proceedings	4 of 34	11.76%	Proceedings	4
Professional Society	2 of 32	6.25%	Professional Society	1 of 34	2.94%	Professional Society	3
Published Book	1 of 32	3%	Published Book	2 of 34	5.88%	Published Book	3
Thesis	2 of 32	6.25%	Thesis	3 of 34	8.82%	Thesis	5
University Website	5 of 32	15.62%	University Website	3 of 34	8.82%	University Website	8
Unknown	0 of 32	0.00%	Unknown	2 of 34	5.88%	Unknown	2
Unpublished Article	4 of 32	12.50%	Unpublished Article	5 of 34	14.71%	Unpublished Article	9
		100%			99.97%		66

\* Totals equal all unique POPs, not total number of theses matched.

# Extent of Matches Found

	Google	Scirus
<b>Entire paragraph or more</b>	<b>8</b>	<b>3</b>
<b>Multiple sentences</b>	<b>5</b>	<b>3</b>
<b>Entire sentence</b>	<b>5</b>	<b>8</b>
<b>Multiple phrases</b>	<b>5</b>	<b>10</b>
<b>Entire phrase</b>	<b>11</b>	<b>8</b>

# Cross-searching results

Unique Theses Matches: Searching Same Phrase in Other Tool

Tool	Total	Found Same	Found Different	Not Found
Scirus	12	2	2	8
Google	14	3	3	8

10 of 68, or 14.71% were missed as a result of the searcher

16 of 68, or 23.53%, were missed as a result of the search engine choice

# Matches by Institution ID

Institution	POPs Found	Theses Checked	Percentage POPs
A	0	1	0%
B	9	10	90%
C	1	3	33.33%
D	0	1	0%
E	0	0	0%
F	1	1	100%
G	6	8	75%
H	2	3	66.67%
I	26	38	68.42%
J	0	3	0%
Total	45	68	66.67%

# Prevalence

- ◆ 27% General Sample of Theses (Study 1)
- ◆ Highest Rates in Sciences & Engineering
- ◆ 67% of Science/Engineering Theses (Study 2)

# Prevalence: Cautions

- ◆ Includes pre- & post-thesis plagiarism
- ◆ Only electronic U.S. master's theses
- ◆ Matches of 7 consecutive words or longer (coincidence?)
- ◆ Difficult to establish originator, collaborative works, etc.


# Prevalence: On the other hand...

- ◆ Limited to word-for-word (verbatim) plagiarism
- ◆ Searching limited to 10 minutes
- ◆ 28% of matches in Study #1 found near beginning of theses (before or on 1<sup>st</sup> page)
- ◆ Unknown what anti-plagiarism methods already applied to theses in sample
- ◆ Only used search engines (not other databases or print resources)
- ◆ If authors (graduate students) of web-published theses are plagiarizing, what about undergraduates?
- ◆ Uncovered several highly probable examples of plagiarism

# Noteworthy Examples (probable cases of plagiarism)

- ◆ Multiple paragraphs in 2003 thesis identical to 2001 doctoral dissertation by different author (no apparent institutional/research affiliation).
- ◆ Entire paragraphs of 2003 thesis identical to chapter in published book (Wiley) from 2001 (no apparent affiliation). Student now appears to be enrolled in doctoral program at same institution where thesis was completed.
- ◆ Opening paragraph (p.1) of thesis is identical to numerous websites
- ◆ Paragraph+ of thesis identical to IEEE article published in 1999 by Australian researchers

# Other observations regarding prevalence

- ◆ Many of the matches seemed less serious– incomplete citations, definitions, etc.
  - ◆ Problem of establishing chronology
  - ◆ Problem of determining collaborative partnerships
- 
- A decorative graphic at the bottom right of the slide, consisting of a stylized mountain range silhouette in shades of teal and blue.

# Implications

- ◆ Students
  - Future members of professional communities
  - Future educators, journal editors
- ◆ Faculty
  - Co-authors with plagiarists?
  - Signing off on theses with plagiarism
  - Students borrowing faculty works
  - Plagiarizing students?

# Implications

## ◆ Institutions

- Research partnerships
- Price of trust and integrity
- Legal issues (Chronicle article)

## ◆ Publishing World

- Anti-plagiarism software purchases
- Price of trust and integrity
- Legal issues, lawsuits

# Suggestions & Solutions

- ◆ Clarity of collaboration
- ◆ Clarity of self-citing
- ◆ Greater indications of previous research
- ◆ Using tools at all levels to instill and/or enforce greater ethics
- ◆ More education of expectations and consequences

# Role of Librarians

- ◆ Web pages, guides for proper citation and plagiarism
- ◆ Librarians offering checking services to faculty
- ◆ Instruction roles: educating faculty and students during library instruction sessions, online tutorials

# Concluding Remarks

- ◆ Using search engines:
  - Select phrases of unique words
  - Conduct phrase searches
  - Look for unquoted statistics, facts, maps, definitions, graphs; changes in language style or vocabulary; undocumented chronologies of research
- ◆ Advantages:
  - Only costs times
  - Quick
  - Could be delegated to staff, student workers
  - Avoids privacy concerns of databases of students' works
- ◆ Disadvantages:
  - Won't catch conceptual plagiarism
  - Shifting content of web: print what you find!
  - Won't find all printed materials

Questions?



# Sources

Presentation charts and content were taken from these sources written by the presenters:

Study 1: Using the Google Search Engine to Detect Word-for-Word Plagiarism in Master's Theses: A Preliminary Study (forthcoming in College Student Journal)

Study 2: Anti-Plagiarism Tools: Scirus v. Google