Plagiarism in publisher files

Steve O’Connor

A talk at the Online Conference,
Sydney February 2005
Talk Outline

• Is plagiarism a real problem?
• Publisher views
• Nature of incidents
• Problem of the intranet
• Open Access Tales of caution
• Software capability
• Legal implications
• Publisher Options
• What to take away
Is Plagiarism a real problem???

• All the research indicates cheating rife; fraud also
• Pressure to publish undiminished
  – RAE exercise impact in UK and emerging in Australia
• Jayson Blair *New York Times*
• Jack Kelley *USA Today*
• Clear evidence in academic journals
• Stephen E. Ambrose and Doris Kearns Goodwin are examples of popular authors caught severely plagiarising others
'AUSTRALIAN art writer Patricia Macdonald yesterday received a personal apology from Robert Hughes for "unconsciously cannibalising" one of her reviews… Hughes, who can still "regurgitate" classical works learnt by rote at school, claims he has been afflicted with a photographic memory since childhood: "It's just something that happens."

*The Australian 7 November 1998*
Information Gathering

Years

%  

Colleagues  
Browsing Journals
Information Gathering

Years

%}

Colleagues
Browsing Journals
Citation Linkage
# Victorian Universities data

## Project summary data

<table>
<thead>
<tr>
<th>Description</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of Subjects (discipline areas)</td>
<td>20</td>
<td>100%</td>
</tr>
<tr>
<td>Subjects with papers with high levels of text matching</td>
<td>14</td>
<td>70.04%</td>
</tr>
<tr>
<td>Total Number of papers</td>
<td>1925</td>
<td>100%</td>
</tr>
<tr>
<td>Paper with non-attributed text matching of: 25% and Greater</td>
<td>166</td>
<td>8.62%</td>
</tr>
<tr>
<td>Attributed text for papers with 5% or greater text matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text non-attributed</td>
<td>269</td>
<td>13.97%</td>
</tr>
</tbody>
</table>
## Extent of copying

<table>
<thead>
<tr>
<th>Volume of Copying of Content of Papers</th>
<th>Number of Papers</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greater than 75%</td>
<td>14</td>
<td>0.73%</td>
</tr>
<tr>
<td>Between 60 - 100%</td>
<td>18</td>
<td>0.94%</td>
</tr>
<tr>
<td>Between 50 - 100%</td>
<td>29</td>
<td>1.57%</td>
</tr>
<tr>
<td>Between 40 - 100%</td>
<td>54</td>
<td>2.81%</td>
</tr>
<tr>
<td>Between 5 - 100%</td>
<td>269</td>
<td>13.97%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1925</td>
<td>100%</td>
</tr>
</tbody>
</table>
Non-attributed text

Essays with non-attributed text as a percentage of whole project

- 86%
- 14%

Legend:
- Red: Non-attributed matched text
- Green: No or little matched text
Characteristics of an average essay judged to be non-attributed

- Non-attributed to other essays: 21%
- Non-attributed web text: 16%
- Attributed text: 2%
- Own work: 61%
Software capability

• To be able to search internet
• To be able to match against intranet files
• Text match
• Search as many other e-resources as feasible
• Most importantly, publishers to protect outcomes
An original paper is as unique as a fingerprint
The digital manuscript is submitted over the Internet.
Checking process

1. Manuscript or article submitted to iThenticate

2. Computer **transforms** manuscript into a digital fingerprint: a very long string of numbers

3. Copy of Internet

   Electronic Books

   Journals / Periodicals

4. Extract matching documents
Generating an Originality Report

Entire process: < 10 seconds

Matching passages from 3+ billion Internet web pages (downloaded at a rate of 40 million pages/day)

Matching passages from thousands of digital books

Matching passages from tens of millions of periodical articles

Compare matching passages to original manuscript or article
Suspect manuscripts are highlighted

<table>
<thead>
<tr>
<th>Author</th>
<th>Title</th>
<th>Report</th>
<th>Links</th>
<th>File</th>
<th>Doc ID</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mickelson, Mallory</td>
<td>Chapters 1-3</td>
<td></td>
<td></td>
<td></td>
<td>1180910</td>
<td>04-15-03</td>
</tr>
<tr>
<td>Mickelson, Mallory</td>
<td>Chapters 4-6</td>
<td></td>
<td></td>
<td></td>
<td>1180910</td>
<td>04-15-03</td>
</tr>
<tr>
<td>Dudson, Frank</td>
<td>Chapters 7-11</td>
<td></td>
<td></td>
<td></td>
<td>1180010</td>
<td>04-15-03</td>
</tr>
</tbody>
</table>
Every instance of matching text is underlined and color-coded indicating the possible source.

Author: Doe, John  
Title: Topics in Neurobiology  
Paper ID: 40657  
Submitted: 03-21-01 12:15 PM

<table>
<thead>
<tr>
<th>Source</th>
<th>Link</th>
<th>Similarity Index</th>
<th>Version</th>
<th>Side-by-Side Version</th>
<th>Match</th>
<th>Reanalyze and Exclude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet</td>
<td><a href="http://www.t2.net/schask/smanual7.html">http://www.t2.net/schask/smanual7.html</a></td>
<td>46%</td>
<td>#1</td>
<td>03-05-03</td>
<td>34%</td>
<td></td>
</tr>
<tr>
<td>Internet</td>
<td><a href="http://www.schoolsucks.com/text/paper.htm">http://www.schoolsucks.com/text/paper.htm</a></td>
<td>13%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet</td>
<td><a href="http://www.psy.plym.ac.uk/year2/schizophrenia/manifest.html">http://www.psy.plym.ac.uk/year2/schizophrenia/manifest.html</a></td>
<td>10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet</td>
<td><a href="http://www.salamon.psy.plym.ac.uk/year2/schizophrenia.html">http://www.salamon.psy.plym.ac.uk/year2/schizophrenia.html</a></td>
<td>8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Database</td>
<td>Submitted to Art Center College of Design on 2002-11-03</td>
<td></td>
<td></td>
<td></td>
<td>3%</td>
<td></td>
</tr>
</tbody>
</table>

Report text:

It is hard to comprehend the pain caused by a chronic disease like schizophrenia without personal experience—either as a victim or by having a relationship with a victim. In actuality, no individual is immune from schizophrenia, it affects individuals, families, communities, and society as a whole. My aunt’s struggle with schizophrenia for twenty years has been challenging for our family, and with a lot of courage she shared her story with me.

It has been twenty-five years since I first became mentally ill. As I approach forty-five, I find myself still struggling with the same symptoms, still haunted by the same fears and paranoia. I am haunted by an evasive picture of what my life could have been, whom I might have become, what I might have accomplished. My schizophrenia is a sad realization, a painful reality that I live with everyday. I probably inherited a predisposition to mental illness, my uncle was diagnosed with dementia praecox, an earlier term for schizophrenia. In my last year in high school, I began to experience personality changes. I did not realize the significance of the changes at the time, and I think others denied them out. Looking back, I can see that they were the earliest signs of the illness. I became increasingly withdrawn and silent. I felt alienated and lonely and hated everyone. I even considered suicide. I felt as if there was a huge gap between me and the rest of the world. Everybody seemed distant from me.

Schizophrenia is defined as a group of psychosis characterized by confused and disconnected thoughts, emotions, and perceptions. Schizophrenia is a brain disorder, which is identified by specific concrete symptoms. Schizophrenia is not a split personality, or multi-personality, and it has been proven that schizophrenia is not caused by childhood trauma, bad parenting, or poverty. Schizophrenia is marked by extreme thought disorder, and is usually treatable with medication. Given proper support, many people with schizophrenia can learn how to deal...
Individual sources can be directly compared to the original manuscript.

It is hard to comprehend the pain caused by a chronic disease like schizophrenia without personal experience—either as a victim or by having a relationship with a victim. In actuality no individual is immune from schizophrenia, it affects individuals, families, communities, and society as a whole. My aunt's struggle with schizophrenia for twenty years has been challenging for our family, and with a lot of courage she shared her story with me.

It has been twenty-five years since I first became mentally ill. As I approached forty-five, I find myself still struggling with the same symptoms, still crippled by the same fears and paranoia. I am haunted by an evasive picture of what my life could have been, whom I might have become, what I might have accomplished. My schizophrenia is a sad realization, a painful reality that I live with everyday. I probably inherited a predisposition to mental illness; my uncle was diagnosed as having dementia praecox, an earlier term for schizophrenia. In my last year in high school, I began to experience personality changes. I did not realize the significance of the changes at the time, and I think others denied them—but looking back I can see that they were the earliest signs of the illness. I became increasingly withdrawn and sullen. I felt alienated and lonely and hated everyone. Even considering suicide, I felt as though there was a huge gap between me and the rest of the world, everybody seemed distant from me.

Schizophrenia is defined as a group of psychoses characterized by

My Two Faces

By Elizabeth Nastan
AP Wire Services

Thirty years ago was when I first became mentally ill. As I approach forty-five, I find myself still struggling with the same symptoms, still crippled by the same fears and paranoia. I am haunted by an evasive picture of what my life could have been, whom I might have become, what I might have accomplished. My schizophrenia is a sad realization, a painful reality that I live with. I probably inherited a predisposition to mental illness; my uncle was diagnosed as having dementia praecox, an earlier term for schizophrenia. In my last year in college, I began to experience personality changes. I did not realize the significance of the changes at the time.
Octavia E. Butler, author of Dawn, Bloodchild, and Parable of the Sower, gives us yet another title to rave about with Parable of the Talents, the last in a series of futuristic communities inevitable of what ours may someday become. Octavia Butler finishes the epic with a tale of the survival, destruction, and rebirth of a community called Acorn with the religion of Earthseed as its foundation. Lauren Olamina’s love is divided among her young daughter, her community, and the revelation that led Lauren to found a new faith that teaches “God Is Change.” However, in the wake of environmental and economic chaos, the U.S. government turns a blind eye to violent bigots who consider the mere existence of a black female leader a threat. Soon Lauren must either sacrifice her child and her followers or forsake the religion that can transform human destiny. The main hero, an African American lady named Lauren Oyina Olamina, mentioned initially in the Parable of the Sower, is the initiator, organizer, and the architect of this secure “heaven” that found it might not shy away from the anarchy of the outer world. Beginning in the not so distant future year of 2032, Olamina and her followers found them in the progressing, successful, profiting shelter of Acorn surrounded by people they love and with whom they share their lives. Yet, they were quick to learn their Earthseed religion of hope and change was not enough to defend their freedom against the invasion of Christian American Crusaders, followers of Jarrett, the new war hungry President. Twenty nine years from now, Lauren has devised a new religion, Earthseed, and formed a small community to live according to its edicts. Lauren Oya Olamina Bankole, to give her full name, wants to expand and spread the belief in Earthseed throughout the world, but has no idea how to go about this.
Publisher Views

• Can make the problem go away
  – Virginia Law Review
  – Elsevier removal of articles from ScienceDirect

• Grossly misleading trails of information especially in digital environment

• Publishers acknowledge quite significant instances

• Pressure on Editors and disciplines to ‘deal with’

• Tendency to push under rug to avoid bad publicity for discipline
Like most journalists, I avoid plagiarism. Taking credit for others' work is abhorrent.

I've never been able to fathom why some writers such as Jayson Blair; formerly of The New York Times, and Jack Kelley, formerly of USA Today, would risk killing their credibility and their gilded career prospects by stealing from other writers or simply making things up.

Imagine my shock, then, when a plagiarism detection program in Oakland, Calif., suggested that some of my work was questionable — possibly dubious.

Checking our work in bytes and pieces

Computer resources help the newspaper investigate the possibility of unoriginal or fabricated reporting.

The gravity of the situation calls for unusual steps such as the Star-Telegram's fact-checking policy.

Checks scrutinize stories in three ways with help from the writers, who provide lists of sources and contact information.

Stacy Garcia, director of news research, compares facts, phrasing and other details with results from searches of the massive LexisNexis database and Web searches. D'Lisa Gibson, Witt's administrative assistant, assists with calls to sources to verify the accuracy of quotes and paraphrases.

Web searches are crucial to the fact-checking process because plagiarists have found it so tempting, easy and relatively safe to steal abroad.

We gain considerable scope, depth and speed with iThenticate.

For instance, in a matter of seconds, iThenticate will make a digital fingerprint of an article and check it against two monolithic collections of published information — the ProQuest database, which is similar to LexisNexis, and an archived Internet database of more than 4.5 billion Web pages.

An additional 40 million Web pages are added to the archive each day as iThenticate searches its Web-based sources for fresh material.

To use the resource, we go to the iThenticate Web site, sign in, paste a story into a box and hit a "submit" button.

We receive an "Originality Report" that gives the per centage of and has been on the Web, but it opens far more of the Web to us — quickly.

I might never have known that 2 percent of my March 28 column about the fact-checking policy could have been challenged as suspicious, or that 5 percent of my April 4 Weekly Review centerpiece on political labeling would have raised an electronic eyebrow.

To my relief, it turned out that iThenticate had found a summary of the column on the Society of Professional Journalists' Web site. Bits of text and quotes in the centerpiece that I'd taken from and attributed to Texas Media Watch were found on that group's Web site.

As Witt told staff members when he announced the fact-checking policy, there's nothing to worry about.
Nature of incidents

- Publisher A publishes monograph which is later discovered to be plagiarised from American author
- Publisher B discovers whole issues of journals pirated
- Publisher C finds article completely plagiarised, withdraws digital copy and advises on print
- Publisher D textbook on Sociology into 4th edition riddled with plagiarism
Problem of the intranet

• Problem for publishers both traditional/commercial and emerging through Open Access

• Value of ‘lost’ reputation ????
• What is the quality checking process internally in an organisation ?
• Intranet goes public
• Institutional repositories
• Open Access movement
Legal issues to face

• Copyright breach
  – Up to US$300,000
  – ‘Harmonisation of Australian and US copyright law
  – Mickey Mouse protection clause
  – Civil to criminal

• Is it better to know or to remain ignorant?
OPTIONS

• Publisher ‘unwillingness’ to see
  – Difficulty alone in keeping up with technology
• Editors cannot be accountable for plagiarism and fraud
  – Peer review is extremely difficult to rely on
• Use of standards such as DOI
  – To identify and to facilitate access to digital articles and graphs, tables
  – Should be used to register ‘originality’ check,
• Publishers to maintain confidentiality with internal scrutiny
Five Principles
Clayton Christensen Innovators dilemma

1. Companies depend on customers and investors for resources
   *Do not meet current customer need fail
2. Small markets do not solve growth needs of large companies
   *Emerging market first mover advantage ..margins too small for large companies..new markets often larger
3. Markets that do not exist cannot be analysed
   *Market research and planning good for sustaining technologies
Five Principles ……. 

4. Organisation’s capabilities define its disabilities
   * Capability in processes and values…work against organisation implementing change

5. Technology supply may not equal market demand
   * Technology improvement provides greater performance than market can absorb
Publishing cycle

User as Author

User as Reader

Librarian as funder

Publisher
Publishing cycle

User as Author

User as Reader

Publisher

DOI
Publishing cycle

User as Author

User as Reader

Publisher

Go direct; pay per view
Digital Object Identifier

• Has the potential to be both:
  – The industry unifier
  – Save authors and publishers with a common descriptor to IP
  – Analogous to MARC standard; except that it can deliver a reader to the object
What can be taken away at end of this conference?

• Acknowledging the IP of others is and will remain a huge problem
• Efforts to address the problem is core for the future of whole industry
• Insurance is better than embarrassment
• DOI key to industry content transportability and originality verification
Excuse me sir, you've been upgraded to pilot.
Steve O’Connor
CEO
CAVAL Collaborative Solutions
steveo@caval.edu.au