



# The CrossRef Experience

Workshop in Shanghai, China

April 2015

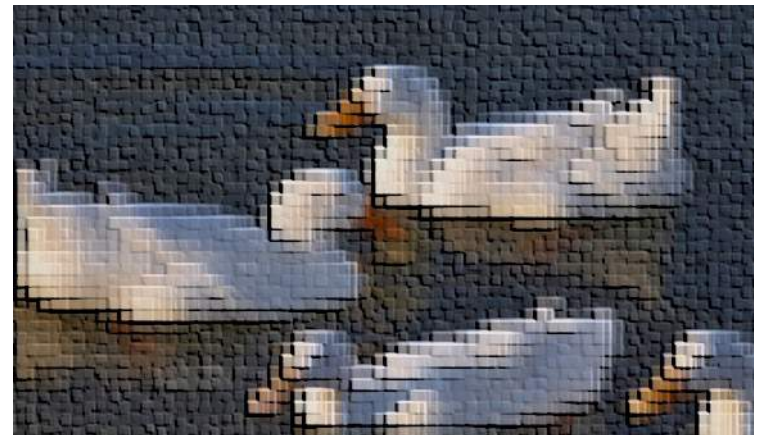
Ed Pentz & Pippa Smart





# CrossCheck

Scanning for plagiarism by detecting duplication



# **THE PROBLEM**

# Plagiarism and unauthorised copying

- Common unethical behaviour:
  - Copying another person's work without attribution
  - Republishing without acknowledgment of the original
- Wastes the time of readers, reviewers, editors
- Fails to give credit to the original authors
- Is fraud !

# Why do we need plagiarism checks?

- International Standards for authors
  - “Researchers should adhere to publication requirements that submitted work is original, is not plagiarised, and has not been published elsewhere.”  
Committee on Publication Ethics (COPE) 2010. [http://  
publicationethics.org/resources/international-standards](http://publicationethics.org/resources/international-standards)
- But
  - There appears to be an increase in plagiarism and duplicate publication

## Why is there an apparent increase in plagiarism?

- Increased pressure to publish (fraud)
- Increasing numbers of researchers without suitable training (ignorance)
- Misplaced respect of other work (ignorance)
- Difficulties in writing in English (or other language) (ignorance)
- Easy to copy another online work (fraud)
- **Better systems to discover plagiarism!**

**THE CROSSREF SOLUTION:  
CROSSCHECK**

# The life of CrossCheck

- 2006:
  - CrossRef board raises plagiarism as area of concern
- Late 2007/early 2008:
  - pilot with seven publishers and technology partner (iParadigms)
- June 2008:
  - CrossCheck launched



# CrossCheck statistics

- 642 publishers
- Over 42 million content items indexed
  - 118,000+ titles
- 190,000+ manuscripts being uploaded every month
- Title and member list on the CrossRef website

# Why not use search engines?

- Why not use Google / Bing / Baidu to check text?
  - Cannot paste in all the text (word limits)
  - Too many results?
  - Cannot always reach the “duplicated” text to investigate fully
    - Perhaps behind a paywall, or archived offline
  - No reporting support

# **HOW CROSSCHECK WORKS**

# What is behind CrossCheck

- iThenticate software (from iParadigms)
  - Good reputation, used by academic institutions
  - Good reporting outputs

# What is checked

- The CrossRef database
  - Members' content
    - more publishers = more content in that database
- The Internet
  - Current and archived pages
- Online and offline databases (via iThenticate)
  - Including large databases from EBSCO, Gale, etc.

# How CrossCheck works

Your document



iThenticate



cross check

arXiv.org

EBSCO HOST

PubMed

GALE CENGAGE Learning

My Documents						Documents	Sharing	Settings	Resubmit	page 1 of 1
Title	Report	Author	Processed	↓	Actions					
Example Article Number One 1 part - 8,274 words	59%	S. Smith	02/09/09							
Example Article Number Three 1 part - 6,983 words	68%	B. Book	02/09/09							
Example Article Number Two 1 part - 3,187 words	7%	J. Jones	02/09/09							

page 1 of 1

Produces a report for each of your documents ...

# Important note !

- Plagiarism is the theft of ideas
  - Reproducing content without attribution
- “Plagiarism-checking” systems cannot identify ideas
  - They identify similar text
    - which might be properly credited
  - They do not identify figures or translated text
    - which might be plagiarised/copied without attribution

# So, remember the limitations

- It won't find images
- Or graphs
- Or tables
- Or formulae
- ... or translations
- ... or ideas
  
- A human must make the final decision



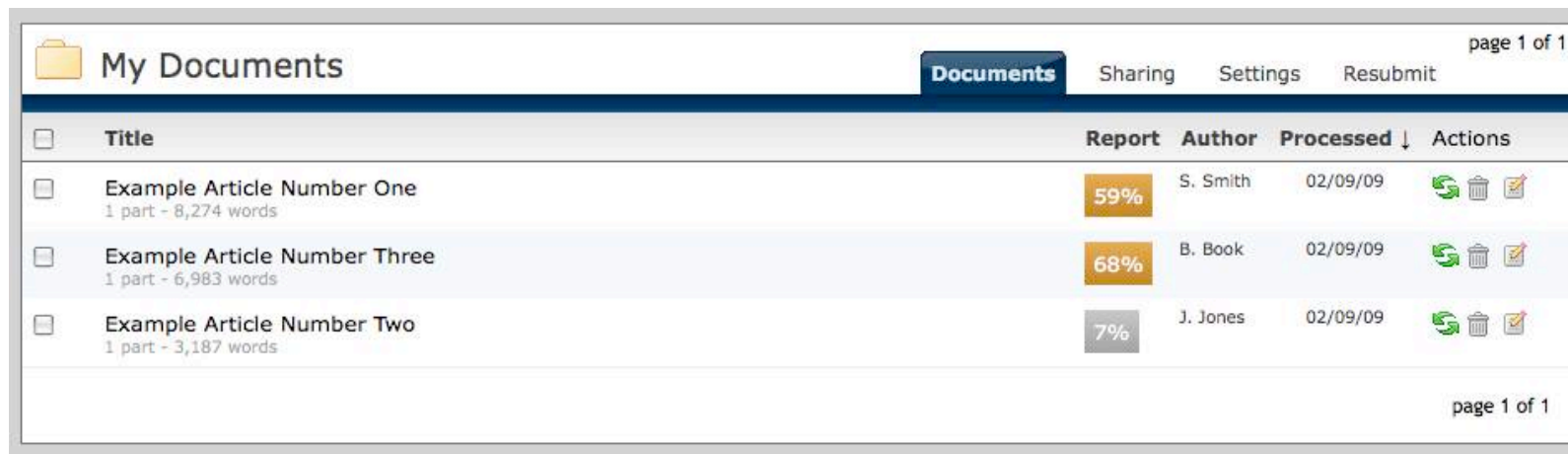
**USING CROSSCHECK**

# How to use CrossCheck ...

- You upload your document
  - You can upload a zip folder (containing several articles)
  - You can upload Word, PDF, HTML, and many other formats
- CrossCheck returns a report
- You investigate
- You make a decision

# What level of match is a problem?

- You set your own levels for investigation
  - 75% 25% 50%
- The percentage is a crude indicator
  - It may be justified reproduction of content



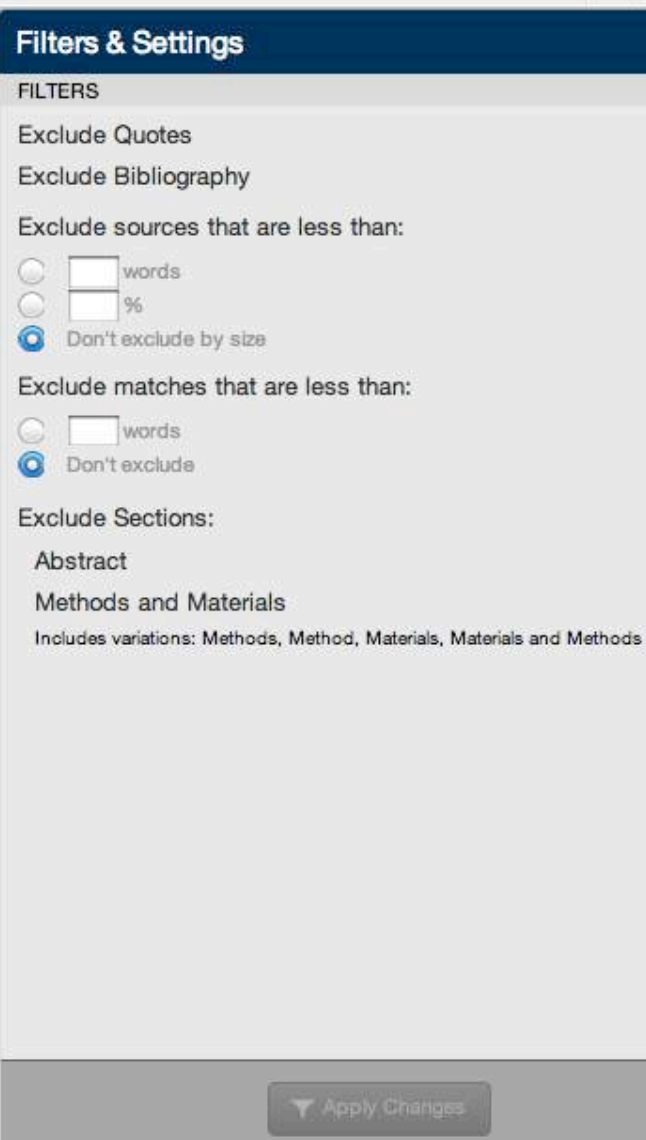
The screenshot shows a web interface for document management. At the top, there is a navigation bar with a folder icon and the text "My Documents". To the right of this are tabs for "Documents", "Sharing", "Settings", and "Resubmit". The main content area is a table with the following columns: "Title", "Report", "Author", "Processed", and "Actions". The table contains three rows of data:

<input type="checkbox"/>	Title	Report	Author	Processed	Actions
<input type="checkbox"/>	Example Article Number One 1 part - 8,274 words	59%	S. Smith	02/09/09	
<input type="checkbox"/>	Example Article Number Three 1 part - 6,983 words	68%	B. Book	02/09/09	
<input type="checkbox"/>	Example Article Number Two 1 part - 3,187 words	7%	J. Jones	02/09/09	

At the bottom right of the interface, it says "page 1 of 1".

# Narrow your searches ...

- You can choose to exclude
  - Quotations
  - Bibliography/references
  - Small sources or matches
  - Sections
    - e.g. exclude the Materials and Methods section



The screenshot shows a 'Filters & Settings' panel with the following options:

- FILTERS**
- Exclude Quotations
- Exclude Bibliography
- Exclude sources that are less than:
  - [ ] words
  - [ ] %
  - Don't exclude by size
- Exclude matches that are less than:
  - [ ] words
  - Don't exclude
- Exclude Sections:
  - Abstract
  - Methods and Materials
  - Includes variations: Methods, Method, Materials, Materials and Methods

At the bottom right, there is a button labeled 'Apply Changes' with a downward-pointing triangle icon.

# Similarity report

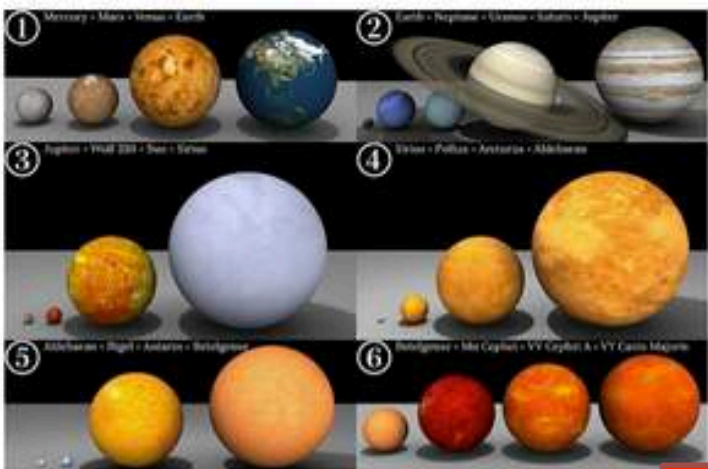
Mar 12, 2012 9:11 AM PST 6,723 words • 230 matches • 3 sources FAQ Paper 1 of 1

**iThenticate** Relation of Mass to Diameter in Neutron Stars Quotes excluded 37%  
Bibliography included 37%

## Relation of Mass to Diameter in Neutron Stars

*B. Asparov*

The combination of the radius and the mass of a star determines the surface gravity. Giant stars have a much lower surface gravity than main sequence stars, while the opposite is the case for degenerate, compact stars such as white dwarfs. The surface gravity can influence the appearance of a star's spectrum, with higher gravity causing a broadening of the absorption lines.



Stars are sometimes grouped by mass based upon their evolutionary behavior as they approach the end of their nuclear fusion lifetimes. low mass stars with masses below 0.5 solar masses do not enter

### Match Overview

1	Internet	25 words	16%
2	Publications	21 words	13%
3	Publications	7 words	8%

**Your article**

**Summary of matches (highest to lowest)**

**You can select a text-only report**

Text-Only Report

When you select the match it shows you where your articles duplicates the source

And you can view the entire source article by clicking here

## 2 Toward a Unified Theory of High-Energy Metaphysics: Silly String Theory

Josiah Carberry  
Department of Psychoce

<http://dx.doi.org/10.5555/1234567>

### Abstract

The characteristic theme of the works of Stone is the bridge between culture and society. Several narratives concerning the fatal flaw, and subsequent dialectic, of semioticist class may be found.

Internet

[psychoceramics.labs.crossref.org](http://psychoceramics.labs.crossref.org)

Journal of Psychoceramics <http://dx.doi.org/10.5555/12345678> Toward a Unified Theory of High-Energy Metaphysics: Silly String Theory Josiah Carberry Department of Psychoceramics, Brown University <http://dx.doi.org/10.5555/12345678> Abstract The characteristic theme of the works of Stone is the bridge between culture and society. Several narratives concerning the fatal flaw, and subsequent dialectic, of semioticist class may be found. T

Full Source View

Note, it uses fuzzy matching – it will pick up word substitutions

This is unique\* to CrossCheck

\*so far as we know

# What fuzzy matching means ...

- Example 1
  - “One possible reason for **these** results is that the ratings reflected **people’s** attitude towards local Government rather than being the result of their interaction with the particular website.”
- Example 2
  - “One possible reason for **the** results is that the **satisfaction** ratings reflected **subjects’** attitude towards **their country’s** local Government rather than being the result of their interaction with the particular website.”

These two paragraphs come up as a match

# When to use CrossCheck

- When do you check articles
  - On submission?
  - On acceptance?
  - Only when you have a suspicion?
- Each journal makes its own decisions
  - Each article costs a checking fee
    - More checking = more expensive
  - Each check (and investigation) takes time
    - More checking = more time



# Within submission systems

- The major submission systems have integrated CrossCheck
  - Articles automatically uploaded/checked when requested
  - If you cannot see it, you need to ask for it to be activated

eJournalPress  
Manuscript Submission

MANUSCRIPT HOME | AUTHOR INSTRUCTIONS | REVIEWER INSTRUCTIONS | HELP | TIPS | LOGIN

CrossCheck Manuscript

200BEJPTST00177 | Mr. Keith Harris | Initial QC Started | The Internal Suicide Debate H

Last Processed: Completed (2009-03-05 10:27:37 by EJP Staff)  
Send File for Processing: [Re-send File](#)  
Similarity Score: 0 %  
View Report: [Display](#)

eJournalPress

Action ▲	Article Type ▲▼	Section/Category ▲▼	Article Title ▲▼
<a href="#">View Submission</a> <a href="#">CrossCheck/Thenticate Results (62%)</a> <a href="#">Details</a> <a href="#">History</a> <a href="#">Technical Check</a> <a href="#">File Inventory</a> <a href="#">Send Back to Author</a> <a href="#">Remove Submission</a> <a href="#">Send E-mail</a>	Research Article		Research Methods in the Physical Sciences

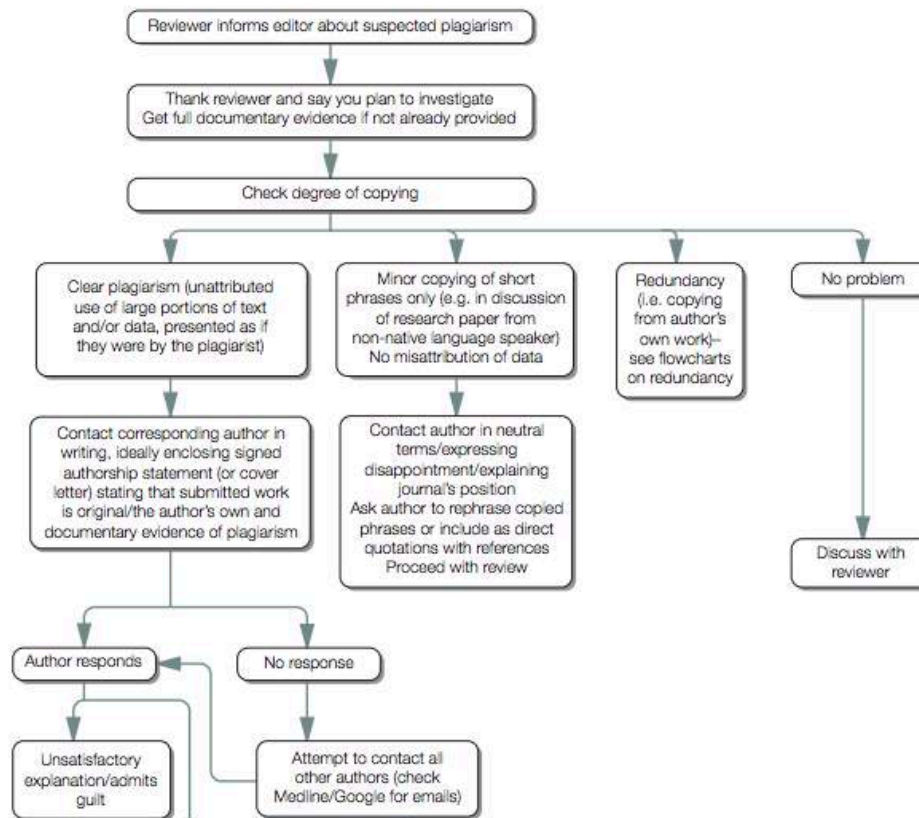
# What happens if you discover plagiarism?

- Refer to the COPE Guidelines and flowcharts
- <http://publicationethics.org/resources/flowcharts>



## What to do if you suspect plagiarism

### (a) Suspected plagiarism in a submitted manuscript



Note: The instructions to authors should include a definition of plagiarism and state the journal's policy on it

# How to avoid plagiarism

- Understand why authors plagiarise
  - Ignorance
    - you need to educate them
  - Fraud
    - You need to respond severely
- Author guidelines
  - Tell authors about your policies on plagiarism
  - Tell authors that you will check articles
- Respond firmly when you discover plagiarism

**WHO CAN USE CROSSCHECK**

# Membership

- You have to be a member of CrossRef
  - But any member can participate in CrossCheck
- CrossCheck fees:
- Two parts:
  - 20% of annual CrossRef membership fee (billed by CrossRef)
  - Fees charged per number of documents you check (billed by iParadigms)
    - From \$0.75 per article
- [http://www.crossref.org/crosscheck\\_fees.html](http://www.crossref.org/crosscheck_fees.html)

# **MAKING THE BEST USE OF CROSSCHECK**

# Participate

- Meetings, user groups, other events
  - User groups at:
    - Council of Science Editors Annual Meeting, Philadelphia, May 15-18
    - ISMTE European Conference, London, October 13
    - CrossRef Annual Meeting, Boston, November 17-18
- Webinars
- Mailing list
- Surveys

# Promotion

- Use your participation to promote the high standards of your journal
- Use CrossCheck to educate
  - Authors
  - Reviewers
  - Editorial Board members