



CrossCheck and Plagiarism Screening: How and what to look for and Future Plans for iThenticate Developments

Rachael Lammey, Product Manager
KAMJE Workshop
Seoul, June 2014



Gyeongju and Seo





CrossRef was founded in 2000 by a group of scholarly publishers for the purpose of establishing a reference linking system using the DOI.



- 24 staf (7 UK, 17 USA)
- 16 person Board of Directors, cross-section of publishers
- Working groups and committees





- 4,208 publishers and societies
- 67,063,793 content items with DOIs
 - 28,582 journals
 - 343,924 books
 - 26,787 conference proceedings



90,000,000

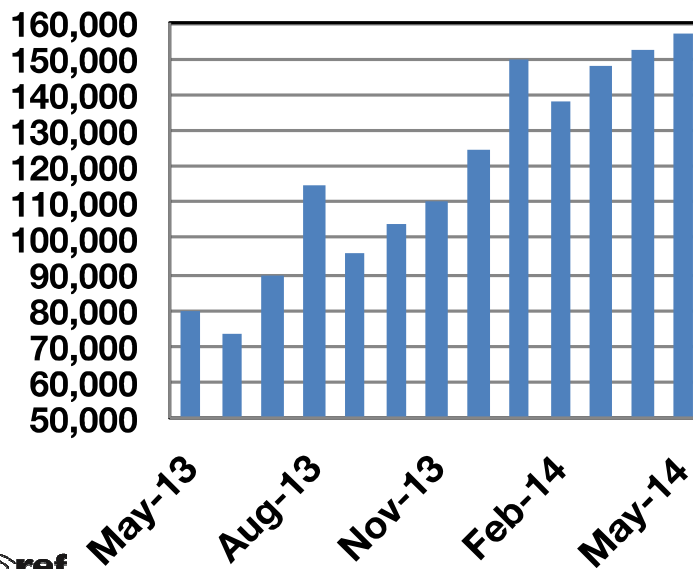




- 572 publishers
- Nearly 40 million content items indexed
- Over 100,000 titles
- Over 150,000 documents being screened each month
- Total documents checked in 2013: 1,098,974 (compared to 629,359 in 2012)
- Total documents checked in 2014 ytd: 744,354 (compared to 382,625 in the same period in 2013)



iThenticate Usage



- The Document Viewer
- File size increase (20MB to 40MB)
- Small match exclusion



- Section Exclusion
 - Ability to exclude sections from a report based on section headings
 - Focus initially on two section headings:
 - Abstract
 - Methods and Materials + variations
 - (i.e. Methods, Materials, Materials and Methods)

Interpreting the Report

- The similarity percentage can be misleading
- Use All Sources view.
- Check for standard description
- Check where the match is from
- Check references in the submitted paper



Common Issues with References

- Not referencing correctly
- Unattributed copying of parts of another's work
- Submitting another's work as your own
- 'Self-plagiarism'/reuse of own work
- Dual submission



Example 2: Referencing

6188 words • 127 matches • 735 sources

<p>Internet - 3 sources 4418 words crawled on 17-Jan-2011 dev.ageing.ox.ac.uk</p>	<p>53%</p>
<p>http://dev.ageing.ox.ac.uk/system/files/workingpaper...</p>	<p>53%</p>
<p>http://dev.ageing.ox.ac.uk/system/files/ageing_horizo...</p>	<p><1%</p>
<p>http://dev.ageing.ox.ac.uk/system/files/5%20harper_...</p>	<p><1%</p>

71-310 292

Internet - 4 sources 211 words
 crawled on 20-Aug-2007
gro-scotland.gov.uk



Example 3: Referencing and report

views
 iThenticate Professional Plagiarism Prevention


Example Article Number Three
 By: B Book
 As of Tue Feb 02, 2010 10:30am GMT
 6,983 words - 1 match - 48 sources

Similarity Index
58%

Mode: [Content Tracking]

Exclude Quotes Include Bibliography Excluding matches < 2% Limit match size

<p>Australian Journal of Emerging Technologies and Society Vol. 5, No. 1, 2007, pp: 1-14 Aesthetics, Visual Appeal, Usability and User Satisfaction: What Do the User's Eyes Tell the User's Brain? Gitte Lindgaard is the Director of the Human Oriented Technology Lab (HOTLab) and a Professor of Psychology, both at Carleton University, Ottawa, Canada. Abstract: The impact of colour on the first impression of a website is discussed in the light of several rather puzzling experimental findings, which suggest that background colour and colour combinations might influence users' subsequent opinion of, and satisfaction with, a site. Theories of and approaches to, studying aesthetics and emotion are outlined briefly. It is concluded that, although the criteria by which people judge visual appeal, user satisfaction and trustworthiness are still unclear, perceived usability appears to be related to the detection of stumbling blocks that hinder smooth interaction with a web site and probably to the orderliness of screens. User satisfaction is a complex construct that incorporates several measurable concepts and is the culmination of the interactive user experience. Experimental results suggest that people may be more satisfied with a beautiful product that performs suboptimally than with a more usable but less appealing product. A glance into the future importance of the topics discussed is offered. Keywords: Aesthetics; emotion; user satisfaction; first impression; mere exposure effect. ©Australian Journal of Emerging Technologies and Society 2007 ISSN 1449-0708 http://www.win.edu.au/ajets Vol. 5, No. 1, 2007, pp: 1-14 Introduction When visiting an art exhibition, why do people dwell for 30 minutes in front of one painting and only 10 seconds in front of another? What is it that makes one experience so compelling - and another barely noticeable, even when two paintings are of the same genre, painted in the same period, in the same style, and by the same artist? I stumbled across this question nearly a decade ago after obtaining some rather puzzling results in a usability test of a local Government web site (Lindgaard 1999). In that evaluation one group of participants rated perceived usability and satisfaction after completing a standard usability test that exposed the worst of the usability problems identified in the preceding heuristic evaluation. Since participants managed to complete only one half of the</p>	<p>6,045 words / 87% - Internet from 14-Mar-2009 12:00AM www.winburne.edu.au [] This source has been manually excluded from the Similarity report.</p> <p>3,675 words / 53% - Internet from 04-Nov-2008 12:00AM hot.carleton.ca [] This source has been manually excluded from the Similarity report.</p> <p>3,675 words / 53% - Internet from 22-Feb-2008 12:00AM hot.carleton.ca [] This source has been manually excluded from the Similarity report.</p> <p>3,674 words / 53% - Internet from 04-Nov-2008 12:00AM hot.carleton.ca [] This source has been manually excluded from the Similarity report.</p> <p>3,671 words / 53% - Internet from 04-Nov-2008 12:00AM www.carleton.ca [] This source has been manually excluded from the Similarity report.</p> <p>3,587 words / 51% - Internet from 01-Mar-2008 12:00AM www.carleton.ca [] This is source #1 in the Similarity report. This source is partially hidden by one or more sources in the Similarity report.</p> <p>679 words / 10% - CrossCheck Lindgaard, G. "What is this evasive beast we call user satisfaction?", <i>Interacting with Computers</i>, 2003(8) [] This source has been manually excluded from the Similarity report.</p>
---	---



Example 3: Referencing and report views

ACM SIG VOI, 3, NOV. 1, 2007, pp. 1-14

anthropological point of view it may seem to be nothing more than sensual attractiveness [Berle's work on arousal; Norman's (2004) notion of "visceral emotion"]. To the legislators of taste it tends to become one aesthetic quality variously differentiated among a number. Those reflecting more generally upon criticism may use it to mean 'aesthetic excellence' [Tractinsky's "expressive aesthetics"]; that is, as an almost empty term, standing for a problem rather than for its solution (Spearshott 1963: 59). [comments and italics added]

Yet, as acknowledged by Spearshott none of these senses of the term correspond to its normal usage, being applied "chiefly to women and weather" (Lindgaard & Whitfield 2001: 375). Aesthetics, like beauty, is thus as elusive as it is confusing. The similarity or overlap between beauty and aesthetics remains undefined; we are unsure about what is being judged (Frohlich 2004), whether they are properties of objects in the world, subjective experiences, emotional reactions residing "in the eye of the beholder", or cognitive judgments (Hassenzahl 2004a;2004b; Norman 2004; Frohlich 2004). Aesthetics therefore lacks an affinity with the main paradigms of psychological and HCI research, and has no secure theoretical attachment point: it lacks a home. Yet, even if aesthetics is a property of objects, when confronted with an object of beauty, it does evoke a positive emotional experience in the viewer.



Internet - 9 sources	4439 words	64%
crawled on 04-Nov-2008 www.carleton.ca		
CrossCheck	576 words	14%
Lindgaard, G., "What is this evasive beast we call user satisfaction?", <i>Interacting with Computers</i> , 200306		
CrossCheck	659 words	9%
Gitte Lindgaard, "Usability and Experiences in the Evolution of Usability", <i>Human-Computer Interaction Series</i> , 2008		
CrossCheck	446 words	6%
Gitte Lindgaard, "Attention web designers: You have 5... milliseconds to make a good first impression!", <i>Behavio</i>		
Internet	426 words	6%
crawled on 30-Sep-2009 anandjlc.hypopad.com		

Example 4: Overlap in specific sections

Procedures

In this experiment, air (an ideal gas) was heated in a pressure vessel with a volume of 1 liter. Attached to this pressure vessel was a pressure transducer and thermocouple to measure the pressure and the temperature, respectively, of the air inside the vessel. Both of these transducers produced voltage.

(K) of the air (the 13.6 psia). In addition, measured pressure was also measured by a pressure transducer. The voltages as measured for the pressure and temperature transducers appear in Table 1. The first two columns of Table 1 show the measured voltages from the pressure transducer and the temperature transducer. Column three shows the measured values of pressures calculated from the

Results and Discussion

This section analyses the results of the experiment. The experiment went as expected with no unusual events that would have introduced error. The voltages as measured for the pressure and temperature transducers appear in Table A-1 of the Appendix. Also included in the Appendix are the equations used for calibrating those voltages with the actual pressures and temperatures. These equations led to the values of pressure and temperature that are shown in the third and fourth columns of Table A-1. From these values, a graph between temperature (K) and pressure (kPa) was created (Figure A-1). As can be seen from the graph, the relationship of temperature versus pressure is roughly linear.

As part of this experiment, the theoretical values of temperature were calculated for each measured pressure value. In this calculation, which used the ideal gas equation, the volume and mass were assumed to be constant. These theoretical values of temperature are shown in the final column of Table A-1. From this final column arose Figure A-2, a graph of ideal temperature (K) versus pressure (kPa). As shown in this graph, the relationship between temperature and pressure is exactly linear.

A comparison between the graph showing measured data (Figure A-1) and the graph showing theoretical data (Figure A-2) reveals differences. In general, the measured values of



Match 1 of 8		
Internet - 2 sources	835 words	100%
crawled on 06-Jan-2009 www.writing.eng.vt.edu		
Internet - 3 sources	376 words	45%
crawled on 19-Nov-2008 www.ece.utah.edu		
Internet - 2 sources	363 words	43%
crawled on 07-Apr-2010 www.usna.edu		
http://www.usna.edu/EE/ee433/Admin/EE433_Project... 43%		
http://www.usna.edu/EE/ee354/Labs/EE354_Lab_Rep... 43%		
Internet	338 words	40%
crawled on 30-Oct-2010 www.cetla.howard.edu		
Internet	82 words	11%
crawled on 30-Jun-2003 www.fbox.vt.edu		
Internet	72 words	9%
crawled on 30-Aug-2012		

Example 5: Potential Problems

Powered by iThenticate

In Google You Trust

Quotes Included
Bibliography Included

67%

by Rachael Lamme

Abstract

Trust and authority are central to academic publishing, but the trust model of the Internet is almost antithetical to the trust model of academia. Publishers have been so preoccupied with the brute mechanics of moving content to the online world that they have virtually ignored the challenge that the Internet trust model poses to the many academic publishers. So it is important that publishers look outside the publishing industry to see what other online industries are doing, as there will be lessons to be learned from that. They can then go forward and re-establish trust in this dynamic environment.

Each day, Internet users are confronted with large amounts of data with no way of being sure it is trustworthy – think about spam emails for example. Indeed, the history of the Internet is full of examples of how those working in technology have underestimated the part that social trust and authority play in communication.

Authority is the key quality of scholarly publishers and librarians, and the trust model

quod.lib.umich.edu

Full Source View

All Sources

Match 4 of 4

Internet - 2 sources 258 words crawled on 07-May-2013 quod.lib.umich.edu	67%
http://quod.lib.umich.edu/jep/3336451.0008.101?rg...	67%
http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jepvie...	67%
Internet - 2 sources 65 words crawled on 02-Sep-2010 www.carl-abrc.ca	17%
Internet - 2 sources 65 words crawled on 06-Jan-2010 devtags.library.upenn.edu	17%
Internet 65 words crawled on 14-Aug-2008 www.adset.org.uk	17%
Internet - 14 sources 65 words crawled on 21-Apr-2011 www.citesonline.org	17%
Internet - 16 sources 65 words crawled on 09-Mar-2012	17%

CrossCheck

CrossCheck as a Deterrent

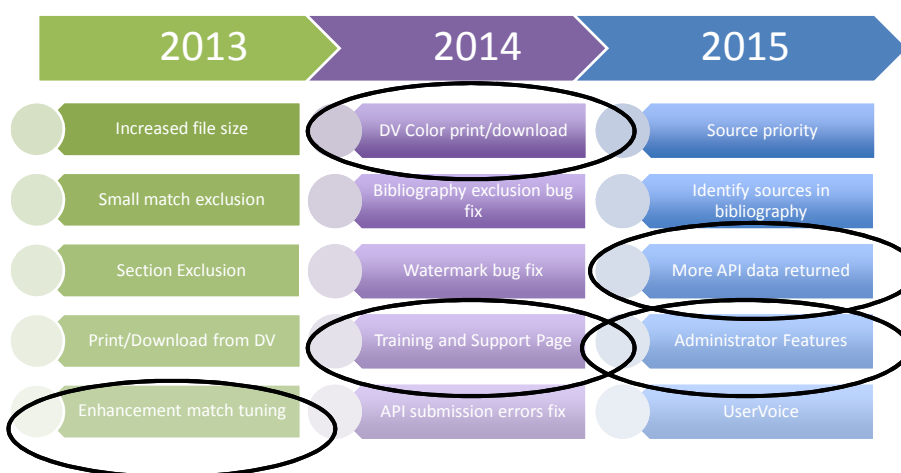
- CrossCheck information on Journal Authors Site and journal homepages
- CrossCheck warnings on peer review systems used by CrossCheck journals

Support from iParadigms

- Free CrossCheck/iThenticate webinars
- On CrossRef website: <http://www.crossref.org/crosscheck/index.html>
- Can do publisher-specific sessions
- Documentation and webcasts
- <http://www.ithenticate.com/resources/customer-training/>
- Information on new functionality
- <http://www.ithenticate.com/products/whats-new/>
- User Groups
- At CrossRef Annual Meeting (workshop day), COPE events, 6th International Integrity and Plagiarism Conference (June 14) and various conferences like CSE
- Email Support
- ccsupport@ithenticate.com
- Support team in UK and in the US



Product Program



- Provide more high level information before investigating a Similarity Report:
 - Display the highest matching source % (already available via API)
 - Display the word count # of the highest match instance
 - MTS development required to include these into their User Interface

Date:	Rev:	Triggered By	File Sent	Report Status	Score	Top source	Highest match	Ignore Score
Mar 18, 2013	0	A Journal	Document A	View	10	1%	30 words	<input type="checkbox"/>
Mar 18, 2013	0	A Journal	Document B	View	20	10%	100 words	<input type="checkbox"/>
Mar 18, 2013	0	A Journal	Document C	View	40	5%	40 words	<input type="checkbox"/>

Example Screenshot



Training and Help Center

CrossCheck Training and Help

Understanding Reports

DOCUMENT VIEWER

[Differences Between Report Modes](#)

[NEW! Understanding Reports in Document Viewer \(video\)](#)

TEXT ONLY

[Differences Between Viewing Modes](#)

Related Reading

[Understanding the Similarity Score \(Article by CrossRef\)](#)

[Manual](#)

[iThenticate Document Viewer Similarity Report Guide \(PDF\)](#)

FAQ

1. What does the Similarity Percentage mean?
2. How do I know if a manuscript has been plagiarized?
3. How do I access the side-by-side comparison mode so that I can compare the manuscript with the original source text?

Known Issues

The list of known issues for the iThenticate are listed on this page:
www.ithenticate.com/known

Create a support ticket

If you have a question or concern that is not answered on this page please contact our support team at csupport@ithenticate.com.



Product Roadmap

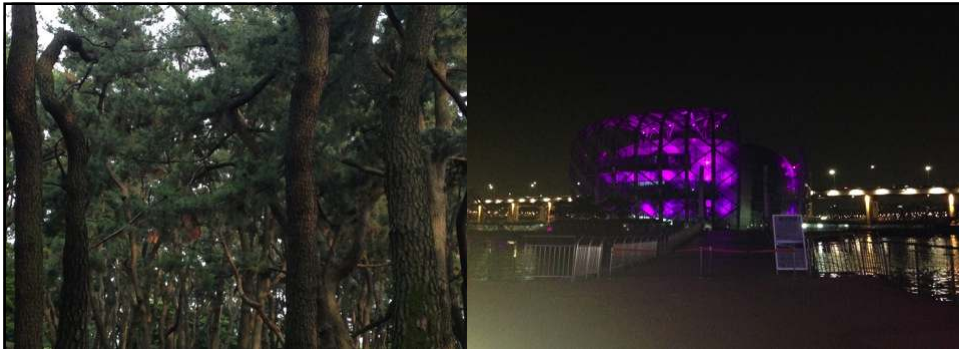
2016+

Image Detection

Support LaTeX

Digital Content Protection

Rejected Manuscript Tracking



Thank You!
감사합니다

<http://www.crossref.org/crosscheck>
rlamney@crossref.org

