

Search Engines and Finding a Threshold Test for Plagiarism

Pheh Hoon LIM [\[1\]](#)

Cite as: Lim P.H., "Search Engines and Finding a Threshold Test for Plagiarism", in European Journal of Law and Technology, Vol 1, Issue 3, 2010.

Abstract

Nowadays, using electronic search engines has replaced visiting libraries for information allowing in-print, out of print and orphan works to be conjured up from cyberspace for ready perusal. Such "allowed" borrowing of content on the internet has created a whole new cut and paste culture making it easy for writers to freely download materials and mix and match during the creativity process. On the other side of the coin, copying and plagiarism, which might have been rife but were more difficult to track in a hard copy world of text and books, are becoming easier to detect. The substantive legal test as to whether copyright has been infringed has not been much affected by the sheer ease of copying digital material; it still remains to be assessed whether a substantial amount from a particular work has been copied, whether the quantum is gauged qualitatively or quantitatively. Plagiarism, however, has never been subject to a comparable substantive test probably because on its own it does not give rise to any legal action. A charge of plagiarism, however, is not without its consequences.

Persons denounced as plagiarists are subjected to sanctions (severe in many academic environments) and social stigma. A recent example is the furore in New Zealand involving Witi Ihimaera's latest published piece of historical fiction, *The Trowenna Sea*. This incident highlighted the finely tuned forensic role that search engines such as Turnitin and Google now play in today's knowledge society in both compiling and unearthing historical memorabilia from all kinds of sources and historical eras. At the same time, every act of reusing a 'free' historical fact or picturesque turn of phrase without fastidious attribution runs the risk of being detected and negatively publicised. It demonstrates the need for a clearer answer for writers, academics and students as to what is acceptable and what is not even when one borrows and builds on material in the public domain. Search engines undeniably play a useful forensic role but have brought to the fore an urgent need for new and clear standards of good practice. Bearing in mind that search engines are heuristic devices, technology aids the plagiarism detection process only to the extent of helping to determine a set of experience-based rules for optimal solutions. While efficient controls are needed at the same time, the process should be thought out on the basis that it remains open to an aggrieved party to pursue a copyright claim if any borrowing

(attributed or otherwise) amounts to a "substantial part" and infringement.

In this paper, the author first explores the nature of plagiarism and its intersection with copyright infringement and then tackles three related issues. The first looks at whether there is a need for a threshold test for plagiarism in terms of the quantum of unacknowledged borrowing that might be permitted before a plagiarism alert is triggered, and whether the test applied could be the same as copyright's infringement threshold, namely substantial taking (whether that be assessed quantitatively or qualitatively). The second is based on the assumption that there should be a standard and questions the appropriateness of allowing technology to nudge the knowledge society towards greater accountability and zero tolerance for borrowing. From a purely pragmatic point of view, were it to become the norm to avoid a claim of plagiarism that all writers must attribute each and every descriptive snippet of information they reuse literary works in future might become overbalanced with more footnotes or endnotes than text. The third issue relates to whether students and academics should be held to a higher standard (effectively zero tolerance) than other writers such as journalists, political speech writers or historical novelists.

1. Introduction

The recent New Zealand furore over Professor Witi Ihimaera's latest published piece of historical fiction, *The Trowenna Sea*, highlighted the highly forensic role that Google (as well as Turnitin software commonly used in academia) now plays in ferreting out plagiarism. The earnest debate this past decade on plagiarism has focused on finding an appropriate definition of the practice and led to the coining of concepts now in common parlance such as "self plagiarism" to denote the recycling of one's own previous work or publications and "facilitated plagiarism" to cover the thriving paper mill industry in which student essays on specific topics are bought and sold. These deliberations have led to the identification of different kinds of plagiarism perpetrated by different groups of writers and resulted in discussions on whether the standard for culpability, should to be fair, be applied variously in these different contexts.

In the current debate over how plagiarism should be defined, some attempted definitions [2] allude to the taking of literary property. Questions have also been raised as to whether the subjective element of intent should be considered, and if so to what extent precisely. From a cultural perspective, claimed lack of intent on the part of an alleged perpetrator may ring true in the case of new foreign students from a rote learning culture. The learning of the Chinese language, for example, starts with a student committing to memory the pictorial strokes of the Chinese characters in order to increase his or her word bank for more advanced reading and comprehension. The ability to learn by rote and memorise is thus essential to academic progress in that culture. Another factor is that some students fresh from other cultures accustomed to treating teachers with considerable reverence might see no reason to paraphrase lecturers' notes or text because in their view the latter have explained theories or subjects so aptly they cannot be improved upon. Indeed Professor Ihimaera when first interviewed by *The Listener* regarding the Victorian texts he had used said "I fell in love with their language and phrasing and I did not feel that I could express their descriptions better!"(a statement he

subsequently qualified when he admitted he had probably done some "unintended" borrowing). [3]

Though access to information is just a mouse-click away, any such borrowing of non-copyrightable ideas from the public domain or copyrightable expression is not spared from scrutiny for plagiarism. Students may plead ignorance, lack of intention or coincidence for failing to measure up to the expectations of academic assessments. Oversight or inadvertence due to the pressure of deadlines and the volume of work are some other reasons given. For academics, "errors" and unintended lapses from otherwise "meticulous practices" may occur. [4]

Well before the advent of search engines for easy retrieval of facts, ideas and information, the early English case of *Pike v Nicholas* [5] found it was natural to use another person's work and that similarities would be found but best practice dictated that sources of information and use should always be acknowledged to avoid allegations of copying. [6] In *Pike*, the idea/expression distinction for copyright subsistence was judicially recognised, [7] with it being accordingly noted that there was "no monopoly in the main theory of the plaintiff" [8] used by the defendant. The plaintiff and defendant had both submitted essays in a contest on a topic relating to the origins of the English nation. Both had researched using common sources and similar theories. In alleging copyright infringement, plagiarised sections were tracked to prove copying. The court denied the copyright claim and stated that where there were common subjects or sources everyone was free to use these for research. [9]

2. Defining Plagiarism

Under copyright principles there is no monopoly in an idea, [10] whereas any definition of plagiarism generally encompasses the borrowing and passing off of ideas as one's own. In New Zealand universities, for example, most definitions revolve around the dishonest use of someone else's work or ideas without acknowledgement with examples of what constitutes plagiarism. [11] Emphasis is placed on the need to properly reference work submitted. [12] The element of intent, which is included in some (but not all) definitions, can be an important factor in determining outcomes as two contrasting Australian decisions demonstrate. In *Re Humzy-Hancock*, [13] the alleged academic misconduct (collaborating and plagiarising work for an assignment and a "take home exam") was made by a student who had sent his soft copy to the other student under the mistaken notion that collaborative work was acceptable practice. The Supreme Court of Queensland, obliged to apply the plain meaning of the definition found in the Griffith Law School Assessment Policy and Procedures, held that intent, which was the "critical question", [14] was absent in the case. As the University had defined plagiarism as the "*knowing presentation*" of the work or property of another person, [15] poor work or the lack of proper referencing was not plagiarism. [16]

This approach differed markedly from that of the Queensland Court of Appeal which had a few years earlier displayed zero tolerance in the case of *Re AJG* [17] when it denied another Griffith University student admission to legal practice for copying the work of another student. In finding academic misconduct the court noted there had been "substantial copying" of another student's work rather than just material taken from the

"public forum". [18] Though there are different views on this contentious point, [19] there is evidence that some institutions are factoring "intent" into the definition rather than viewing plagiarism as a strict liability issue. The Curtin University of Technology and the University of New South Wales in Australia, for example, identify plagiarism by providing for three levels of severity, assessing severity according to the element of intent. [20] In New Zealand the element of intent is showing up in more recent searches of some websites. [21]

3. Contrasting Copyright Infringement and Plagiarism

Copyright and plagiarism are two contrasting but sometimes interrelated concepts. Copyright infringement is a legal wrong in reaping where one has not sown, [22] whereas plagiarism is the unethical taking of another's idea or work without acknowledgement, concepts which clearly intersect but remain distinct legally and ethically. Copyright protects only original works but not the unoriginal facts or ideas contained in the work. It allows the taking of ideas and free use of facts from the public domain. The copying of any such unoriginal part of the work will not amount to an infringement as copyright protects only the expression of the idea expressed in a tangible form. [23] It is no defence, however, to prove an innocent mind or unconscious taking if infringement occurs. [24]

On the other hand, any borrowing (including copyrighted expression), if duly attributed, is not plagiarism. While no plagiarism or copyright issue arises when only ideas and words are taken and attributed, such attribution does not avoid an infringement claim in cases of substantial borrowing of copyrighted expression. In Professor Ihimaera's case where there was no concurrent claim of copyright infringement, the furore waged over his lack of attribution was, nevertheless, very intense.

3.1 The Intersection between Copyright Infringement and Plagiarism

While plagiarism is not necessarily copyright infringement and vice versa, the determination of copyright infringement usually starts with the issue of plagiarised work, as in the case of *Pike* above. In *Harman Pictures N.V. v Osborne* [25] attempts to prove copying also commenced with a comparison of plagiarised phrases between the two works in question. The plaintiffs had accused the defendants and film-makers that the film "Charge of the Light Brigade" had infringed the copyright in their book *The Reason Why*. Phrases from the film script such as "*arranged signal*" and "*a triumph my Lord*" were compared to phrases in the book which used "*the signal*" and "*the review was a triumph for the lieutenant-colonel*" respectively. Both the film and book employed the words "*snooks*" and "*guffaw*". [26] Though the similarities arose through common sources and unoriginal facts relating to some "well known event in history", [27] Goff J was sufficiently impressed by the "marked similarities in the choice of incidents" and the "juxtaposition of ideas" from a copyright view to allow an interlocutory injunction against the defendants on appeal. [28]

In contrast, Mummery LJ in the Court of Appeal in the *Da Vinci* case [29] dismissed as trifling in amount and importance [30] nine instances of such tracking that had been intended to show "word for word similarity" existed between the two books at issue in the

case. The claimants' book *The Holy Blood and the Holy Grail*, purportedly a piece of historical conjecture, [31] revolved around a central theme based on the early history of the Crusade, Knights Templar and the Merovingian kings. Attempting to prove that Dan Brown, author of *The Da Vinci Code*, had infringed the copyright in their book the claimants tracked plagiarised phrases such as "*Jesus survived*", "*Langdon reveals*" and "*Constantine*" in three research documents and "*Jesus the Man*" used by Brown. [32]

Mummery LJ held, however, that unlike plagiarism, copyright infringement was concerned not so much with "language copying" and did not necessitate the tracking of actual language or similar words copied. The emphasis was on whether there had been a substantial copying of an "original collection, selection, arrangement and structure" of literary facts and information. [33] It was stressed as well that no one could monopolise historical research. While the assortment of historical facts and information no doubt needed time and money; [34] protection, however, did not necessarily extend to the ideas used. As he put it: [35]

"The literary copyright exists in HBHG by reason of the skill and labour expended by the claimants in the original composition and production of it and the original manner or form of expression [...] It does not, however, extend to clothing information, facts, ideas, theories and themes with exclusive property rights, so as to enable the claimants *to monopolise historical research* or knowledge and prevent the legitimate use of historical and biographical material, [...]. (emphasis added)"

This same point was made in the 1978 case of *Alexander v. Haley* [36] a copyright infringement case concerning Alex Haley's moving epic and historical novel *Roots* tracing his African ancestry. [37] Frankel D J denied the plaintiff her copyright infringement claim against Haley on the basis that: [38]

"[The Plaintiff] claimed similarities are based on matters of historical or contemporary fact. No claim of copyright protection can arise from the fact that plaintiff has written about such historical and factual items, even if we were to assume that Haley was alerted to the facts in question by reading Jubilee."

He went on to state: [39]

"Another major category of items consists of material traceable to common sources, the public domain, or folk custom. [...] Where common sources exist for the alleged similarities, or the material that is similar is otherwise not original with the plaintiff, there is no infringement. This group of asserted infringements can no more be the subject of copyright protection than the use of a date or the name of a president or a more conventional piece of historical information."

Though both the above decisions concerned claims of copyright infringement, the following quote from the judge in *Alexander v. Haley* does point up precisely how difficult it can be to avoid the borrowing and use of certain words, language or phrases: [40]

"Yet another group of alleged infringements is best described as clichéd

language, metaphors and the very words of which the language is constructed. Words and metaphors are not subject to copyright protection; nor are phrases and expressions conveying an idea that can only be, or is typically, expressed in a limited number of stereotyped fashions."

Similarities do arise as common ideas might only be capable of being expressed in a stereotyped fashion. [41] While no copyright attaches to such use of historical or contemporary facts, the ethical need to acknowledge the source of information remains. It might be time to acknowledge that it would be an encumbrance to require pedantic attribution of every borrowing that can typically be expressed only in a limited number of ways. Such recognition calls for new standards of practice and controls to help formulate guidelines in an internet environment that "allows" instant access to an ever increasing bank of information in digital form.

4. Substantial Copying

4.1 Copyright Infringement in Terms of a "Substantial Taking"

Substantial similarity in copyright cases is based on a comparison of the two works in dispute and a determination whether plagiarised portions are infringing. Proof of similarity based on access to the work leans towards a prima facie presumption of copying, [42] but no actionable infringement arises from coincidental similarities or use of same sources. [43] Once the first issue of actual copying has been established, the second issue of substantiality is then considered. [44] Whether "the whole or a substantial part" of an original work has been copied [45] can be assessed quantitatively or qualitatively. Though a "substantial part" is not defined under copyright legislation, quality more so than quantity goes towards determining infringement. [46] A way to test this might be based on whether the part taken is "novel or striking or merely a commonplace arrangement of ordinary words". [47]

A comparison for substantial similarity is most easily conducted in cases involving literary works where the extent of copying can be tracked based on evidence of "slavish or verbatim word for word" copying, the focus being on the "selection, arrangement and structure" of the works. [48] On the other hand, if a case does not involve literal copying, making it difficult to track "identifiable language", substantial similarity may still be assessed through parallel incidents and materials lifted by the plagiarist. [49]

4.2 Substantial Copying and Plagiarism

Worthy of note here is that in the academic plagiarism case *Re AJG* mentioned above, the Court focused on the fact that there had been "substantial copying" of another's work in assessing the extent of the student's misconduct. [50] Current and newer guidelines drawn up by some universities determine whether a case of plagiarism is "significant" or otherwise by querying the "extent or amount" of plagiarised material, [51] terms that suggest assessing what has been taken in terms of quantity and quality as well. Similarities between two works are easily quantified by the Turnitin mechanism in

percentage figures. Perhaps not surprisingly, those charged with conducting the forensic process tend to be influenced by how substantial the copying is based on these figures in the Turnitin originality reports.

The next step, rightly or wrongly, queries the quantum of unacknowledged borrowing that may be permitted before it triggers an alert for investigative purposes. [52] The reality of the exercise for investigators is to start skimming through the percentages for a cut-off point. There is a strong presumption of plagiarism when the quantity of similar work is substantial, as in higher percentages revealed. However, a lower overall similarity index of below 15 or above 20 per cent in a Turnitin report might be deemed unacceptable if it reveals two substantial paragraphs uplifted verbatim from another source. On the other hand, the 15 per cent may reveal similarities only in generic words or jargon and phrases scattered throughout a paper, without a clear pattern presumptive of guilt. This usually happens when there is a common assignment and a large group of students are all referring to a list of recommended readings or common sources.

Professor Ihimaera stated by way of defence that the copied or unacknowledged passages brought up by a Google search amounted to only 0.4 per cent of his 528-page novel. [53] However, though the percentage did not appear substantial quantitatively, the quality of what was taken based on a visual search proved to be a critical factor. As the example below illustrates, the almost verbatim reproduction of four lines of distinctive prose reflects substantial taking, quality wise: [54]

"Mark the poverty-stricken mother standing in a doorway with a pale-faced child in her arms. Listen to the sounds that proceed from the wretched-looking home with the broken windows: they are the everyday noises of a father swearing in his drink and ragged children crying for their supper."- p49, *The Trowenna Sea*, by Witi Ihimaera

"Mark that poverty-stricken mother who is standing at yonder door with the pale-faced child in her arms [...] Listen to the sounds which proceed from the wretched-looking house with the broken windows; they are the everyday noises of a father swearing in his drink, and children crying for their supper." - "The Manufacturing Poor", by Robert Lamb, *Fraser's Magazine for Town and Country*, VOI XXXVII, January 1848 (As cited in *The Victorian Novelist: Social Problems and Social Change*, by Kate Flint)

As with copyright, it appears that the substantial similarity enquiry process in alleged plagiarism situations is likely to be more swayed towards finding culpability where the quantum borrowed is more qualitatively than quantitatively assessed.

5. Search Engines and Academic Honesty

5.1 The Role of Search Engines

The use of the Turnitin mechanism as a plagiarism detection tool by many tertiary institutions reflects the increasing concern underlying the ethical conduct of students and

the standard of work submitted. The Turnitin tool assists the forensic process of detecting student copying by trawling through the archived papers in its database, which helps towards determining a set of experience-based rules for optimal solution of cases. Ironically, Turnitin had to surmount a copyright infringement challenge itself in accessing its own database content for comparison purposes. [55]

Besides Turnitin, the Google search engine plays a role as well, highlighting similarities outside the Turnitin database of archived materials. Search engines play a useful role making access a less time consuming and laborious task compared to the hard copy world of texts and books. Not all similarities, however, should be of concern as typical words and phrases could be highlighted and linked to many sources. One good example would be copyright's own peculiar phrase "There is no copyright in an idea, only in its expression". A Google search by the author provided more than five million of not very meaningful hits in 0.30 seconds. [56] Another search for the two words "idea expression" came up with more than 20 million hits in 0.20 seconds. [57]

Such burgeoning databases of information can compromise the effectiveness of a search, when the original source is lost and the number of matches can continue to a point which is not quite useful. [58] The Turnitin system works on colour alerts to show the amount of similarities regardless of context, substance or expression in terms of percentages. The figures are subject to further investigation and a visual search with no focus on a copyright idea expression divide. A student may be accused of having plagiarised from another student when both are quoting the same sources or recommended readings. Another main weakness of the Turnitin device is that it does not identify plagiarism by distinguishing between similarities that are properly referenced and similarities that are not.

5.2 Finding a Threshold Test for Plagiarism.

In contrast to the legal threshold for copyright infringement, plagiarism guards against unethical borrowing that is not duly acknowledged. With its idea expression distinction, ideas in copyright fall outside the substantiality enquiry for infringement purposes. Applying a "copyright substantiality" analysis to plagiarised words, ideas and expression might appear problematic at first blush as plagiarism does not require evidence of substantial taking even for ideas. As any unattributed borrowing (whether a little or substantial) makes the plagiarist liable, a zero tolerance approach tends to influence the investigative process, though in practice it rarely pans out (and should not) as the most conclusive or determinative factor in final decisions made in plagiarism cases.

It is definitely difficult to conclude on a quantum of unacknowledged borrowing permitted, as reliance on figures presented by search engines would not provide optimal solutions. This is so as the similarities (quantity wise) based on percentages captured by Turnitin or the number of hits by Google may raise false alerts or false positives. Furthermore, with the element of intent included in the definition of plagiarism by some institutions, unconscious plagiarism would be analysed differently from intentional copying or the "knowing presentation" of another's work. The lack of intent would affect the analysis of a quantitative taking, even if substantial. Qualitatively, a lesser taking of another's work might still be viewed as "significant".

A visual search certainly assists the qualitative assessment bearing in mind the fact that similarities cannot be merely based on unattributed word for word tracking as well, as they may be typical stereotyped words and ideas. The exercise itself is more an art than a science and the use of search engines, despite their forensic potential, offers mainly a heuristic method of problem solving. In view of the inherent limitations in depending on electronic searches, copyright's substantiality threshold to assess plagiarism more so based on quality rather than quantity offers good guidance for maintaining academic rigour.

5.3 Zero Tolerance in the Context of Search Engines

Technology is helping shape the knowledge society towards more ethical boundaries as search engines such as Google Book Search [59] allow more access to content. In a rising culture of digital borrowing, zero tolerance would arguably deter the taking of short cuts and the passing off of others' work as one's own. However, guidance in line with that for copyright infringement analysis, viz to avoid trifling comparisons, [60] must be borne in mind too.

In pragmatic terms, zero tolerance should not be interpreted to mean that every mention made or turn of phrase, proverb or idiom used needs to be referenced. As search engines provide instant access and nudge students, authors and researchers towards more ethical conduct, it should not lead to a plethora of plagiarism alerts for each and every descriptive snippet of information reused. Such confusion was not unknown in the initial implementation of the Turnitin system as a plagiarism detection tool when a few lines of similarities were queried, causing undue concern. [61] The adoption of zero tolerance as a standard throws up the spectre whereby authors must attribute every little turn of phrase used to avoid plagiarism allegations against them. Opponents of the test might well argue that literary works would become overbalanced with more footnotes or endnotes than actual text. While the question of how much borrowing is permitted before it triggers a plagiarism alert is a difficult one, some norm must be set leaving it to time to crystallise out more clearly what amounts to a *de minimis* taking.

For the legal test in copyright infringement analysis, the lack of intent is no defence, albeit ideas are spared from scrutiny. Including an element of intent into the definition of plagiarism would allow for some leeway for unconscious plagiarism or inevitable borrowing of stereotyped words and phrases to be analysed differently from intentional copying or the "knowing presentation" of another's work. Where such copying intersects into the expression of the idea, it would take culpability from the ethical realm into copyright realm for infringement.

6. The Standard to be Applied

6.1 Students

Exploiting another's ideas or facts for better grades in academic work is fundamentally wrong. Students are being penalised in the education process for non-ethical conduct through grade penalties, assisted by search engines such as the Turnitin mechanism.

Admittedly, students and academics are held to a higher standard (namely zero tolerance) shaped by traditional standards of scholarly pursuits in order to preserve the credibility and reputation of their institutions.

In terms of self plagiarism, there could be the occasional student who might not fully grasp the concept possibly due to ignorance, especially when they are repeating a paper. However, academic standards do dictate that students should not use work previously submitted for assessments and guidelines require signed declarations to this effect. Such submissions are easily detected by the Turnitin mechanism through its database of archived papers. Educational institutions now view it essential to identify such cases and approach it under the teaching and learning process as a step towards inculcating ethical conduct and maintaining standards.

Interestingly, a thought provoking point brought up during a European conference discussion suggests that plagiarism could be condoned if such borrowing allows creativity and original work to flourish. A contentious point, obviously, it is worthwhile nevertheless to contemplate ways to veer away from a singular focus on plagiarised work and allow credit for the ways in which the learning outcomes have been met. Based on the approach in some Australian universities and the experience in New Zealand institutions, interest is now being shown in distinguishing between "cut and paste", "lack of attribution" or "poor referencing" breaches and more serious breaches involving dishonesty and cheating such as essays bought from other sources. [62] Poor work not written substantially by students should be assessed under the marking criteria. Termed Level 1 plagiarism (not considered academic misconduct) to Level 3 plagiarism (serious misconduct) guidelines and procedures for staff dealing with student plagiarism such as those by the Curtin University of Technology and the University of New South Wales in Australia treat the former as part of the teaching and learning process under the marking criteria with prevention foremost in mind, while Level 3 plagiarism would require different sanctions under a disciplinary code. [63]

6.2 Academics

The same rule of attribution to guard against plagiarism applies to all students and staff in academic institutions. [64] Academics, though not subject to a formal grading process, can be subjected to just as great or even greater scrutiny. Indeed standards appear to be very high for all senior university personnel as the vice chancellor of Griffith University found out to his cost in 2008 when it was negatively publicised that he himself had plagiarised from Wikipedia for an opinion piece he wrote for a newspaper. [65] The focus on academic honesty can hold academics to a higher standard than other writers such as political speech writers or historical novelists. Indeed had Witi Ihimaera not been a Professor of English and Creative Writing as well as one of New Zealand's most well known authors, the furore over his lack of attribution might not have been nearly as intense.

In relation to self plagiarism and academic recycling or cannibalism of already published work, there is "little consensus" as to whether it can seriously be categorised as being "intellectually dishonest", [66] although another argument sometimes proffered is that the absence of attribution when quoting from one's published work is "genuine plagiarism" as it "deceives" its audience into believing that it is original work. [67] Plagiarists act

unethically by passing off others' work as their own but of course this element of passing off is absent when a work belongs to oneself. A pedantic approach to best practice to allow more transparency would nevertheless require all self attributions to be made.

6.3 Other Writers

Borrowing in the literary world of novels and historical fiction is not an academic exercise. Take, for example, Witi Ihimaera's book *The Trowenna Sea* which is historical fiction rather than a piece of research by an academic historian. The unattributed snippets on Maori rituals, that it includes according to Karen Sinclair (author of *Prophetic Histories: The People of Maramatanga*) [68] belong to the "descendants along the Whanganui River" and need no citation to the book. [69] Even presupposing that the true origin of such snippets can be traced, the point remains that even the first to research some aspect of history do not own the facts and information, meaning it is freely available for use and reuse by other writers of historical fiction.

Even so, in Ihimaera's case, it is hard to ignore sixteen nuggets of good old fashioned prose reproduced almost verbatim. Extreme cases of verbatim or "slavish" copying will be obvious and changes that have been made slightly do not easily disguise plagiarism. [70] It would thus not be advisable or necessary to adopt a strict literal approach for every unattributed picturesque turn of phrase, making technology an encumbrance on the creative and learning process. As Ihimaera's novel proves, the borrowing of others' ideas and historical facts does create for society's benefit a new work of art and literature which many will enjoy. Professor Laurence H Tribe of Harvard University faced concerns raised regarding the lack of acknowledgement of information taken for his 1985 book, *God Save This Honorable Court*. In his statement, he had said that it was a "well meaning effort to write a book accessible to a lay audience through the omission of footnotes or endnotes" which he differentiated from his normal practice for his "scholarly writing". [71]

Newspapers and magazines, though, are not consumed by large sections of footnotes and endnotes, and while blatant copying and reproduction of sections of other journalists' work are subject to disapproval, the loss of reputation does not quite parallel that in academia in terms of grade penalties, suspension or possible expulsion. Arguably, the codes of conduct for news agencies and publishing houses differ from mainstream academia and scholarly pursuits. [72] In the case of Maureen Dowd (Pulitzer prize winner journalist of the New York Times caught also in plagiarism's mire) a formal correction in the next day's paper seemed to settle the matter. The offending almost verbatim paragraph as set out below: [73]

"More and more the timeline is raising the question of why, if the torture was to prevent terrorist attacks, it seemed to happen mainly during the period when *the Bush crowd* was looking for what was essentially political information to justify the invasion of Iraq."

was "inadvertently" taken from another blogger J Marshall:

"More and more the timeline is raising the question of why, if the torture was to prevent terrorist attacks, it seemed to happen mainly during the period when

we were looking for what was essentially political information to justify the invasion of Iraq."

The novel *Roots* by Haley in *Alexander v. Haley*, [74] could lead readers to think that the book offers historical facts regarding the author's very own ancestry and roots. In fact, however, its main character Kunta Kinte turned out to be a non-existent ancestor after the plagiarism was ferreted out. Nevertheless, his moving epic against the backdrop of African history and slavery was enjoyed by many others without any expectations of academic rigour. Besides the copyright infringement claim, plagiarism allegations brought by another party were settled for a sum of money. [75] Thus, whether Haley ought to be held to higher standards (or in Ihimaera's case those parts that the "iwi" [76] found no fault with can be left undisturbed) might best be between him and his readers. Search engines do play a useful forensic role but the process need not be taken out of context and proportion as it remains open to an aggrieved party to pursue a copyright claim if any borrowing (attributed or otherwise) amounts to a "substantial part" and infringement.

A recent case involving Helen Hegemann's prize winning novel "*Axolotl Roadkill*" has created publicity over yet another bestseller with plagiarised passages taken from a blog. Her statement in defence was that "There's no such thing as originality anyway, just authenticity." [77] Despite the plagiarism charges the jury awarded a book prize. A look at the published comparison of a particular passage from her book to the blog [78] did not quite reveal slavish copying but showed, arguably, borrowing of stereotype language in the particular genre.

7. Conclusion

Inasmuch as the quantity and quality of taking do affect both the analysis of copyright infringement and plagiarism, it is difficult to determine an allowed quantum of unacknowledged borrowing before a plagiarism alert is triggered. With the advent of Google Book Search, the internet is poised to deliver much more in terms of search and access. However, reliance on figures revealing the quantity of copying could lead to too many false alerts. As Turnitin process itself reveals, the percentage figures revealed by the overall similarity index and individual matches do not determine whether a student has plagiarised. [79]

In finding a "threshold" test for plagiarism, copyright's infringement threshold in terms of a substantial taking based on quality more so than quantity is often applied in practice to determine liability. The search engine era brings to the fore a need for new standards of good practice and efficient controls to be set up at the same time. The various initiatives discussed in this paper that have been taken by some academic institutions to identify and accommodate different levels of plagiarism appear to be going some way to meeting the need to modify existing protocols and put new formulations in place.

In arriving at an appropriate standard to maintain academic rigour and ensure greater accountability the simple rule of meticulous attribution has much to commend it. Indeed, cultural differences in perception aside, the need to be vigilant and fastidious as to the documentation of sources has always been a part of academia. The advent of increasing sophisticated technology is a constant and pressing reminder that a zero tolerance

approach would be good practice academically to avoid embarrassment. However ideal though, it is not realistic to predicate this rule on meeting the goal of a zero per cent match by search engines or null similarities in a visual search. Search engines do not provide a science for plagiarism detection nor are the percentage figures and visual searches a consummate determinant of what an acceptable threshold can be.

Notes

[1] Auckland University of Technology, New Zealand.

[2] Developed by the United States Legal Writing Institute (defines plagiarism in terms of "the passing off of others' literary property as one's own work without acknowledgement") at <http://lwionline.org/publications/plagiarism/policy.pdf> (retrieved 29 Dec 2009).

[3] Guy Somerset "The incredible likeness of being" *New Zealand Listener* November 14-20 2009, 17. ("The Listener")

[4] *Ibid* , at 15.

[5] (1869-70) L.R. 5 Ch. App. 251.

[6] *Ibid*, at 269.

[7] See P Samuelson "Why Copyright Law Excludes Systems and Processes from the Scope of Its Protection" (2007) 85 Texas Law Review 1921, 1925. Note that predating *Pike*, another case *Millar v Taylor* (1769) WL 17 (KB) stated too that "sentiments are free and open to all; and many people may have the same ideas upon the same subject" (at p 2358).

[8] *Pike*, *supra* note 5 at 268.

[9] *Ibid*, at 259-261.

[10] *Designers Guild Ltd v Russell Williams (Textiles) Ltd* [2000] 1 W.L.R. 2416, at 2420 per Lord Hoffman citing *Morritt LJ* in the Court of Appeal [2000] F.S.R. 121 at [37].

[11] AUT University 2010 Calendar , p 6 at http://www.aut.ac.nz/_data/assets/pdf_file/0010/99469/academic-calendar-2010.pdf .

See also the websites for:

University of Auckland (<http://www.auckland.ac.nz/uoa/home/about/teaching-learning/honesty/tl-about-academic-honesty>);

University of Canterbury (<http://library.canterbury.ac.nz/services/ref/plagiarism.shtml#1>);

University of Otago (<http://www.otago.ac.nz/study/plagiarism/otago006308.html>)

(retrieved 2 and 4 March 2010).

[12] A more detailed analysis of the definitions used in New Zealand universities was discussed in the author's paper presented to the 18th Annual Conference of the Australia

and New Zealand Education Law Association (ANZELA), Melbourne, Australia, 30 Sep - 2 Oct 2009.

[13] [2007] QSC 34.

[14] *Ibid* , para. 14.

[15] *Ibid*.

[16] *Ibid* , paras. 18 & 30.

[17] [2004] QCA 88 (Unreported, de Jersey CJ, Jerrard JA and Philippides J, 15 March 2004).

[18] *Ibid* .

[19] See for example L Corbin and J Carter "*Is Plagiarism Indicative of Prospective Legal Practice?*" (2007) 17 Legal Education Review 53, 54 (preferring a strict liability approach).

[20] See "Plagiarism Policy and Procedures" Curtin University of Technology at <http://academicintegrity.curtin.edu.au/local/docs/staffguide.pdf> and "Procedures for Dealing with Student Plagiarism - Handbook for Staff (Revised June 2006)" University of New South Wales (UNSW), Australia at http://www.lc.unsw.edu.au/plagiarism/Procedures_Student_Plagiarism_2006.pdf (retrieved 15 February 2010).

[21] See websites for the University of Otago, *supra* note 11 and the University of Auckland at

<http://www.auckland.ac.nz/webdav/site/central/shared/about/teaching-and-learning/academic-honesty-and-plagiarism/about-academic-honesty/academic-honesty-brochure-final.pdf> (retrieved 4 March 2010).

[22] *Designers Guild* , *supra* note 9, at 2418 per Lord Bingham of Cornhill.

[23] See discussion of *Pike* and *Millar*, *supra* note 6. (The idea expression distinction has since been codified under [17 U.S.C.A. § 102\(b\)](#) and given recognition under Article 9(2) of the World Trade Organisation TRIPS Agreement.

[24] *Baigent v The Random House Group Ltd* [2007] EWCA Civ 247, para. 106.

[25] [1967] 1 W.L.R. 723.

[26] *Ibid* , at 734 -736.

[27] *Ibid* , at 727.

[28] *Ibid* , at 735.

[29] *Baigent* , *supra* note 24.

[30] *Ibid*, para. 148.

[31] *Ibid*, para. 20.

[32] *Baigent v Random House Group Ltd* [2006] EWHC 719, paras. 321-323.

[33] *Baigent* , *supra* note 23, para. 145.

[34] *Ibid*, paras. 154 -155.

[35] *Ibid*, para. 156.

[36] [1978] 460 F.Supp. 40.

[37] Alex Haley *Roots* (Pan Books Ltd, Great Britain, 1978).

[38] *Alexander, supra* note 35, at 44-45.

[39] *Ibid*, at 45.

[40] *Ibid*, at 46.

[41] See also *Reyher v Children's Television Workshop* [1976] 533 F.2d 87, 91.

[42] *Francis Day & Hunter Ltd. v Bron* [1963] Ch 587, 612.

[43] *Baigent supra* note 23, para. 117.

[44] *Ibid*.

[45] Section 29(2)(a) Copyright Act 1994 (New Zealand).

[46] *Ladbroke (Football) Ltd v William Hill (Football) Ltd* [1964] 1 W.L.R 273, per Lord Reid at 276.

[47] *Ibid*.

[48] *Baigent , supra* note 23, paras. 141 & 145.

[49] *Sheldon v Metro-Goldwyn Pictures Corporation* (1936) 81 F.2d 49, 54-56 (per Learned Hand J).

[50] *Re AJG, supra* note 17.

[51] See "Procedures for Dealing with Student Plagiarism - Handbook for Staff " UNSW, *supra* note 20.

[52] First year students caught out by the Turnitin system for plagiarised work constantly query (during the interview process with the author) whether 25 per cent or thereabouts was a threshold that was condoned. Apparently, it was a figure plucked from the student grapevine!

[53] Jared Savage "Plagiarists 'like drug cheats'" *The New Zealand Herald* (Auckland, New Zealand, November 20 2009) A3.

[54] *The Listener , supra* note 3, at 16.

[55] *A.V. v iParadigms* (2009) 562 F.3d 630. (In the quest to detect plagiarism, Turnitin's practice of comparing the contents of student papers against its ever burgeoning database of student papers led to the challenge that Turnitin's owners were infringing copyright themselves).

[56] Google came up with 5,260,000 in 0.30 seconds at <http://www.google.co.nz/#hl=en&source=hp&q=There+is+no+copyright+in+an+idea+%2C+only+in+its+expression&btnG=Google+Search&meta=&aq=f&oq=&fp=f9df6c68fbb c2a74> (retrieved 1 Jan 2010).

[57] Google came up with 20,100,000 hits for "idea expression" in 0.20 seconds at <http://www.google.co.nz/search?hl=en&q=+idea+expression&btnG=Search&meta=&aq=f&oq> = (retrieved 1 Jan 2010).

[58] A Turnitin report can show an overall similarity index of 63 per cent comprising a 31 per cent match between two reports and the remaining 32 per cent scattered in one or two per cent similarities to numerous other sources.

[59] See Complaint, *Author's Guild v Google Inc.*, No. 05 CV 8136 (S.D.N.Y. filed 20 Sep 2005) and Complaint, *McGraw-Hill v Google Inc.*, No. 05 CV 8881 (S.D.N.Y.

filed 19 Oct 2005) and the resulting proposed class settlement agreement. (Google's controversial efforts to create a massive digital database of books and texts in the public domain as well liaison with a public and four university libraries will enable consumers to search texts of books online).

[60] See *supra* note 30.

[61] The author's previous experience advocating for students at the University of Auckland, New Zealand, where the Turnitin plagiarism mechanism was used.

[62] See V Goldblatt "The Perils of Plagiarism: Processes for Managing Academic Misconduct" Refereed Proceedings of the ANZELA 18th Annual Conference (see *supra* note 12 for details of the conference). The University of Otago, New Zealand, distinguishes between (and provides guidelines for determining) two levels of offences at <http://www.otago.ac.nz/administration/policies/otago003145.html> (retrieved 4 March 2010).

[63] See websites for Curtin University of Technology and UNSW, *supra* note 19.

[64] *The Listener*, *supra* note 3, at 18 (Ihimaera's comment: "There is not one rule for students and another for university staff").

[65] Michael Sainsbury "Uni Chief Lifted Text from Wikipedia" *The Australian* April 26, 2008 at <http://www.theaustralian.com.au/news/uni-chief-lifted-text-from-wikipedia/story-e6frg6oo-1111116167447> (at 23 Dec 2009) (on the discussion and reference to the doctrines of "Unitarism" and "Wahhabism").

[66] See C M Bast and L B Samuels "Plagiarism and Legal Scholarship in the Age of Information Sharing: The Need for Intellectual Honesty" (2008) 57 *Cath. U. L. Rev.* 777, 784-787.

[67] S P Green "Plagiarism, Norms and the Limits of Theft Law: Some Observations on the Use of Criminal Sanctions in Enforcing Intellectual Property Rights" (2002) 54 *Hastings L. J.*

167, 191.

[68] Karen Sinclair *Prophetic Histories: The People of Maramatanga* (Bridget Williams Books Ltd, New Zealand, 2002).

[69] *The Listener*, *supra* note 2, at 18.

[70] *Baigent*, *supra* note 23, para. 141 (per Mummery LJ).

[71] See Harvard Plagiarism Archive "The Tribe Transgression: Professor Tribe's Statement of September 26, 2004" (the material was taken from Professor Henry Abraham's work regarding the political history of Supreme Court appointments) at http://authorskeptics.blogspot.com/2005/04/tribe-transgression-professor-tribes_22.html (retrieved 28 Dec 2009).

[72] See also Green *supra* note 66, at 199.

[73] Marcus Baram "Maureen Dowd Admits Inadvertently Lifting Line from TPM's Josh Marshall", 17 May 2009 at http://www.huffingtonpost.com/2009/05/17/maureen-dowd-admits-inadv_n_204418.html (retrieved 27 Dec 2009).

[74] See *supra*, note 36.

[75] "The Celebrated Roots of a Lie" at <http://www.martinlutherking.org/roots.html> (retrieved 27 Dec 2009).

[76] New Zealand Maori nomenclature referring to the tribe - see Sinclair, *supra* note 67, at 229 (Glossary). (Ihimaera had consulted the iwi according to protocol - *The Listener*, *supra* note 3, at 18).

[77] Nicholas Kulish "Not Plagiarism but Mixing and Matching, says Best Selling German Author, 17" *The New York Times* February 11, 2010 at <http://www.nytimes.com/2010/02/12/world/europe/12germany.html?hpw> (retrieved 12 Feb 2010).

[78] See details of the passage in Tony Paterson "Publish and be damned: Young writer's ego dramatically punctured" *The Independent Friday, 19 February 2010* at <http://www.independent.co.uk/arts-entertainment/books/news/publish-and-be-damned-young-writers-ego-dramatically-punctured-1904037.html> (retrieved 19 Feb 2010).

[79] Turnitin "Key Questions Students Ask" at http://www.turnitin.com/resources/documentation/turnitin/sales/Turnitin_Questions_Students_Ask.pdf (retrieved 23 Dec 2009).