

UNDERSTANDING AND AVOIDING PLAGIARISM

By User:Awadewit, User:Elcobbola, User:Jbmurray, User:Kablammo,
User:Moonriddengirl and User:Tony1 , 13 April 2009

Plagiarism, as Wikipedia's article on the topic explains, "is the use or close imitation of the language and ideas of another author and representation of them as one's own original work." At best it is intellectual sloppiness and at worst outright theft.^[1] As Robin Levin Penslar notes in *Research Ethics: Cases and Materials*, "The real penalty for plagiarism is the abhorrence of the community of scholars."^[2] It can bring a community into disrepute. Wikipedia's editors should create their own articles, not adopt the work of others. But while this is an easy approach to recommend, plagiarism may not be as simple as it first seems—it is often committed inadvertently. The best way to prevent plagiarism is to understand clearly what it is, how to avoid it, and how to address it when it appears.

UNDERSTANDING PLAGIARISM

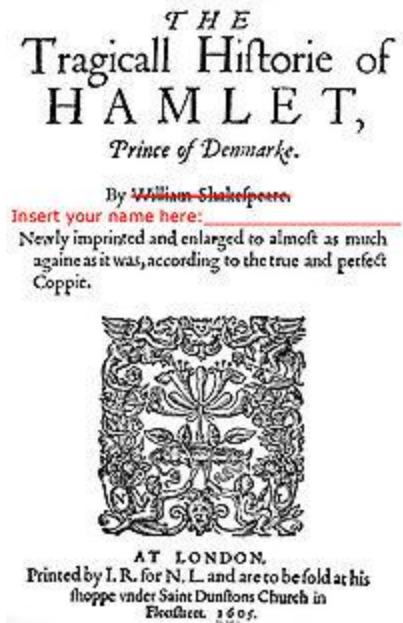
Wikipedia is not a primary source and contains no original research; therefore, everything that appears on Wikipedia should be rooted in a reliable source. The problem with plagiarism is not that it involves the use of other people's ideas, but rather that other people's words or ideas are *misrepresented*—specifically that they are presented as though they were "an editor's own original work". Even if contributors provide a citation for a sentence, it may still be plagiarism if they do not clearly indicate with quotation marks the duplication of the source's wording. Citations are universally understood as indicating a source for information, not as a license to copy the original wording.

There are three major ways to plagiarize:

1. Failing to acknowledge the source of quotations and borrowed ideas;
2. Failing to clearly mark copied language with quotation marks;
3. Failing to sufficiently adapt a summary or paraphrase and thus following the wording of a source too closely^[3]

PLAGIARISM AND COPYRIGHT INFRINGEMENT

Plagiarism is not the same as copyright infringement: material can be plagiarized from both copyrighted and public domain sources.^[4] One report about a plagiarism scandal on Wikipedia claimed that "Wikipedia editors ... declared a handful [of the allegedly plagiarized articles] to be OK because copied passages came from the public domain."^[5] If this was indeed the reaction of Wikipedia editors, they were mistaken. To clarify this, think of the famous opening line of Jane Austen's novel *Pride and Prejudice* (1813): "It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife."^[6] The text of this novel, like the text of the 1911 *Encyclopedia Britannica*, is in the public domain. However, these are Austen's words and even though no one owns the copyright to them any longer, we need to acknowledge that the wording is hers. By inserting this sentence without quotation marks into an article, Wikipedia editors would be plagiarizing Austen.^[7] Apart from the ethical need to credit her for her words, Wikipedia has a scholarly duty to inform its readers of the source of such a sentence, including the page number where the sentence can be found in the source.



Wikipedia policies say much about copyright violation, but far less about plagiarism. The guideline on the topic was written only in 2008. However, Wikipedia's co-founder Jimmy Wales took a clear stand on the issue in 2005: "Let me say quite firmly that for me, the legal issues [surrounding plagiarism] are important, but far far far more important are the moral issues. We want to be able, all of us, to point at Wikipedia and say: we made it ourselves, fair and square."^[8]

WHAT TO CITE: THE "COMMON KNOWLEDGE" EXCEPTION

Not every fact contained in a Wikipedia article *requires* attribution. When a fact is "common knowledge"—that is, generally known—it is not plagiarism to repeat it, even if contributors learned it from a specific reference. For example, it is commonly known that Emily Dickinson published very few poems during her lifetime.^[9] Generally, if information is mentioned in many sources, especially general reference sources, and easily found, it is considered common knowledge. It is also acceptable to reproduce non-creative lists of basic information, such as an alphabetical directory of actors appearing in a film. While Wikipedia's verifiability policy encourages the citing of such information, a failure to do so is *not* plagiarism.

Although common knowledge and non-creative lists of basic facts do not "belong" to a source and do not require attribution to avoid plagiarism, less commonly known information, opinions and creative text *do*. Likewise, the creative *presentation* even of common knowledge, belongs to its original author. Contributors can safely re-use the fact, but *not* the language unless it is a title, as for a job or a creative work, or utterly devoid of creativity, such as a common phrase. From a

copyright standpoint, the level of creativity required to claim ownership is minimal. The United States Supreme Court has indicated that under US copyright law, which governs copyright matters on Wikipedia, "[t]he vast majority of works make the grade quite easily, as they possess some creative spark, 'no matter how crude, humble or obvious' it might be."^[10] Similarly, most text will be creative enough that its replication will be plagiarism. Accordingly, while text such as "Dickinson was born on December 10, 1830" can be copied without quotation marks, care must be taken not to rely too much on the presumption that text is not creative. Further, one cannot copy an entire source in this way, claiming that it is "common knowledge" or uncreative text. In such cases, it can come down to the length of a string of exactly copied words; good editors get a feel for where it's starting to be dishonest not to attribute.

Less commonly known facts or interpretations of facts must be cited to avoid plagiarism, and creative text must either be quoted or properly revised.

AVOIDING PLAGIARISM

To construct articles that read smoothly while still remaining faithful to their sources, it is essential to learn how to properly use other people's ideas and words. Wikipedia contributors need to know when to give credit, how to adapt source material so that it can be used in an article, and when to use quotations.

QUOTATION

When editors want to use verbatim excerpts of a source, there is one simple way to avoid plagiarism: use direct quotations. The words from the source should be reproduced exactly as they appear in the original, enclosed within quotation marks, and identified by an inline citation after the quotation. However, direct quotations should not be overused. They run the risk of copyright infringement if the sources used are not free. Wikipedia's non-free content guidelines offer some guidance on when to use direct quotations and remind us that the "[e]xtensive quotation of copyrighted text is prohibited." But even when free sources are used, the overuse of direct quotation produces articles that are simply collections of quotations. The risk is a fragmentary effect in which the broader context of the quoted material is unclear, and readers are left to piece together the information, which often involves shifts in writing style.

Quotations should generally be used in the following situations:

- "When language is especially vivid or expressive"^[12]
- "When exact wording is needed for technical accuracy"^[12]
- "When it is important to let the debaters of an issue explain their positions in their own words"^[12]
- "When the words of an important authority lend weight to an argument"^[12]
- "When the language of a source is the topic of your discussion"^[12]

ADAPTING SOURCES: PARAPHRASING AND SUMMARIZING

Source text is usually adapted using a combination of paraphrase and summary. These two styles generally differ in their level of detail. A summary is more likely to be used for longer expanses of text and to cover only the major points in a passage, omitting or touching lightly on examples or definitions; a summary is generally expected to be considerably shorter than the original source. By contrast, paraphrasing is more likely to be closer to the original and may be nearly as long as or even longer than the source.

Adapting source text, whether by paraphrasing or summarizing, is a valuable skill, and contributors to Wikipedia need to be alert to the potential for inadvertent plagiarism. Many editors believe that by changing a few words here or there—or even by changing a great number of the words found in the original source—they have avoided plagiarism. This is not necessarily the case. Nor does the mere rearrangement of clauses, sentences, or paragraphs avoid the problem.

Problems in paraphrasing

In this example, Wikipedia's article text is an attempt at paraphrasing the source. However, almost all of the original word choice, word order and sentence structure is retained.

Source

"A statement from the receiver, David Carson of Deloitte, confirmed that 480 of the 670 employees have been made redundant ... At least 100 Waterford Crystal employees are refusing to leave the visitors' gallery at the factory tonight and are staging an unofficial sit-in. The employees say they will not be leaving until they meet with Mr Carson. There were some scuffles at one point and a main door to the visitors' centre was damaged ... Local Sinn Féin Councillor Joe Kelly, who is one of those currently occupying the visitors' gallery, said the receiver had told staff he would not close the company while there were interested investors."

Wikipedia article:

"A statement issued by the receiver, Deloitte's David Carson, confirmed that, of the 670 employees, 480 of them would be laid off. The workers responded angrily to this unexpected decision and at least 100 of them began an unofficial sit-in in the visitors' gallery at the factory that night. **They insisted they would refuse to leave until they had met with Carson.** Following the revelations, **there was a minor scuffle during which the main door to the visitors' centre was damaged.** Local Sinn Féin Councillor Joe Kelly was amongst those who occupied the visitors' gallery."

Analysis:

- "A statement issued by the receiver, Deloitte's David Carson, confirmed that, of the 670 employees, 480 of them would be laid off" vs. "A statement from the receiver, David Carson of Deloitte, confirmed that 480 of the 670 employees have been made redundant". – The structure of Wikipedia's statement is essentially the same as the original. Changing a single word and slightly reordering one phrase is not enough to constitute a paraphrase.
- "They insisted they would refuse to leave until they had met with Carson" vs. "The employees say they will not be leaving until they meet with Mr Carson". – The structure of this sentence is the same.
- "there was a minor scuffle during which the main door to the visitors' centre was damaged" vs. "There were some scuffles at one point and a main door to the visitors' centre was damaged". – The structure and language of the two sentences are the same.
- "Local Sinn Féin Councillor Joe Kelly was amongst those who occupied the visitors' gallery" vs. "Local Sinn Féin Councillor Joe Kelly, who is one of those currently occupying the visitors' gallery". – This slight rewording does not change the fact that the underlying structure and language are the same. Minor changes, such as "was amongst those" --> "is one of those" and "occupied" --> "currently occupying", are not enough to constitute an original rewriting of the passage.

GOOD ADAPTATION PRACTICE

In terms of both plagiarism and copyright, the author of a text not only "owns" the precise, creative language he or she uses, but less tangible creative features of presentation, which may incorporate the structure of the piece and the choice of facts. In terms of plagiarism, but not copyright, the author *also* "owns" the facts or his or her interpretation of them, unless these are, as mentioned above, common knowledge. Revising to avoid plagiarism means *completely restructuring a source in word choice and arrangement while giving due credit for the ideas and information taken from it*.

In this paraphrase, the language and structure of the passage has been significantly altered, making it an original expression of the ideas. The ideas have, of course, been properly credited.

Source:

"In earlier times, surveillance was limited to the information that a supervisor could observe and record firsthand and to primitive counting devices. In the computer age surveillance can be instantaneous, unblinking, cheap, and, maybe most importantly, easy." — From Carol Botan and Mihaela Vorvoreanu, "What do Employees Think about Electronic Surveillance at Work"? p. 126^[13]

Paraphrase:

"Scholars Carl Botan and Mihaela Vorvoreanu claim that the nature of workplace surveillance has changed over time. Before the arrival of computers, managers could collect only small amounts of information about their employees based on what they saw or heard. However, because computers are now standard workplace technology, employers can monitor employees efficiently (126)."^[13]

This adaptation, from the featured article about Thomas Eakins' *The Swimming Hole*, displays attribution of opinion and uses a combination of paraphrase and quotation:

The Swimming Hole represented the full range of Eakins' techniques and academic principles. He used life study, photography, wax studies, and landscape sketches to produce a work that manifested his interest in the human form.^[14] Lloyd Goodrich (1897–1987) believed the work was "Eakins's most masterful use of the nude", with the solidly conceived figures perfectly integrated into the landscape, an image of subtle tonal construction and one of the artist's "richest pieces of painting".^[15] Another biographer, William Innes Homer (b. 1929), was more reserved and described the poses of the figures as rigidly academic. Homer found inconsistencies in paint quality and atmospheric effect, and wrote that the painting was unsuccessful in reconciling antique and naturalistic ideals. For him, "it is as though these nudes had been abruptly transplanted from the studio into nature".^[16]

Unfortunately, there is no hard and fast rule for how much revision is necessary to avoid plagiarizing. In evaluating *copyright* concerns, the United States courts adopt a "substantial similarity" test that compares the pattern and sequence of two works, finding such similarity where "the ordinary observer [reading two works], unless he set out to detect the disparities, would be disposed to overlook them, and regard their aesthetic appeal as the same."^[17] Even if all of the language is revised, a court may find copyright infringement under the doctrine of "comprehensive non-literal similarity" if "the pattern or sequence of the two works is similar".^[18] Likewise, plagiarism may exist if readers comparing the two works would come away with a sense that one is copied from or too heavily based on another.

Editors should always compare their final drafts with the sources they have used to make sure that they have not accidentally come too close in language and structure or failed to attribute when necessary.

RESEARCH AND WRITING METHODS: TIPS FOR AVOIDING PLAGIARISM



One way editors can minimize the tendency to reuse text is to *not* copy and paste text into their working drafts. Instead, editors should assemble and organize their notes, excerpts, and other source materials by topic. This can be done either in hard copy or by using an electronic filing system. Editors should then read and absorb what the sources say and proceed to writing a draft version, in their own words, of each topic. These drafts can be assembled according to the editor's own organizational schema. There are a number of ways to organize material; editors should not slavishly follow a source's structure, either in overall organization, or in the composition and arrangement of sentences and paragraphs within each section. This method reduces the temptation (and makes it harder) to adopt verbatim language and organization from the sources.

At the same time, when taking notes from a source for their own use, editors may find it useful to take them verbatim, with quotation marks, if they will not have access to that source as they are writing their final draft. If a different language is used in note-taking, an editor may find him or herself accidentally restoring some of the author's original words when constructing a draft. Being able to see at a glance exactly how the source was written can help avoid this.

Use multiple sources, if possible. Editors may find it more difficult to avoid following that text too closely if they rely on only one source, as they will necessarily be limited to those details selected by the author of that original source. It is not impossible to revise and reorganize a single source sufficiently to avoid plagiarism or copyright infringement, but it is more difficult.

Notes

1. ^ Leight, David (1999). "Plagiarism as metaphor". In Buranen, Lise and Roy, Alice Myers. *Perspectives on Plagiarism and Intellectual Property in a Postmodern World*. Buffalo: SUNY Press. p. 221. ISBN 0791440796.
2. ^ Penslar, Robin Levin (1995). *Research ethics: cases and materials*. Bloomington: Indiana University Press. p. 148. ISBN 0253209064.
3. ^ Hacker, Diana. *A Pocket Style Manual* (5 ed.). Boston: Bedford/St. Martin's. p. 107. ISBN 0312559933.
4. ^ Fishman, Stephen (2008). *Public domain: how to find & use copyright-free writings, music, art & more* (4th, illustrated revised ed.). Nolo. p. 35. ISBN 1413308589. "To avoid charges of plagiarism, authors of scholarly works ... always give proper credit to the sources of their ideas and facts, as well as any words they borrow. This is so even if the work borrowed from is in the public domain."
5. ^ Jesdanun, Anick (4 November 2006). "Wikipedia Critic Finds Copied Passages". *The Sydney Morning Herald*. Associated Press.
<http://www.smh.com.au/news/breaking-news/plagiarism-uncovered-on-wikipedia/2006/11/04/1162340080487.html>. Retrieved 6 April 2009.

6. ^ Austen, Jane (1914) [First published 1813]. *Pride and Prejudice*. Chicago: Charles Scribner's Sons. p. 1. <http://books.google.com/books?id=s1gVAAAAAYAAJ>. Retrieved 18 March 2009.
7. ^ The requirement for the use of quotation marks on Wikipedia to delineate duplicated public domain text, where attribution is obvious, is disputed. See Wikipedia talk:Wikipedia Signpost/2009-04-13/Dispatches.
8. ^ Wales, Jimmy (28 December 2005). "Comment". Wikipedia. http://en.wikipedia.org/w/index.php?title=Wikipedia:Administrators%27_noticeboard/Incidents&diff=prev&oldid=33005781. Retrieved 31 March 2009.
9. ^ Hacker offers this as an example of common knowledge (p. 107).
10. ^ *Feist Publications v. Rural Telephone Service*, 499 U.S. 340 (United States Supreme Court, 1991).
11. ^ Truss, Lynne (2004). *Eats, Shoots & Leaves: The Zero Tolerance Approach to Punctuation*. New York: Gotham. pp. 151–52. ISBN 1592400876.
12. ^ ^{a b c d e} Hacker, p. 110.
13. ^ ^{a b} Hacker, p. 109.
14. ^ Sewell, Darrel; Kathleen A. Foster, Philadelphia Museum of Art, Musée d'Orsay, Metropolitan Museum of Art (2001). *Thomas Eakins*. New Haven: Yale University Press. p. 100. ISBN 0300091117.
15. ^ Goodrich, Lloyd (1982). *Thomas Eakins, Volume I*. Cambridge: Harvard University Press. pp. 239–40. ISBN 0674884906.
16. ^ Homer, William Innes (1992). *Thomas Eakins: His Life and Work*. New York: Abbeville. p. 116. ISBN 1-55859-281-4.
17. ^ *Peter Pan Fabrics, Inc. v. Martin Weiner Corp.*, 274 F.2d 487, 489 (2d Cir., 1960).
18. ^ *Arica v. Palmer*, 970 F.2d 106 (2d Cir., 1992).