

Self-plagiarism in Computer Science

By Christian Collberg
and
Stephen Kobourov

*April 2005 Communications of the ACM,
Volume 48 Issue*

Presented by Miaomiao Zhang

Summary

This article uses some good examples to explain how often self-plagiarism happens and how it might be defined.

It develops some terminology to describe actions which might be referred to as self-plagiarism.

Good points of this article:

This article mentioned a sensitive topic in computer science and give us a rough definition and analysis of self-plagiarism. And hope us to realize the importance of it, because it has the following bad effects on our research community:

- It can give the public the idea that research dollars are spent on old results, rather than on original one;
- It can indicate to us that academic dishonesty is not a big problem and this could lead to more serious forms of academic dishonesty becoming more acceptable;
- Author who rewrite their results again and again get more rewards, those who publish each result only once get less;
- A self-plagiarized paper is allowed to be published, while more deserving paper is not.

What actions might be referred to as self-plagiarism?

Textual reuse: text/ images

Semantic reuse: ideas

Blatant reuse: texts or ideas (the two works are virtually indistinguishable)

Selective reuse: bits and pieces (parts)

Incidental reuse: texts or ideas (not directly related to the new ideas)

Reuse by cryptomnesia: texts or ideas (unaware of the existence of that work)

Opaque reuse: texts or ideas (not acknowledging the existence of that work)

Advocacy reuse: texts or ideas (different community)

After author introduced above terminology, he mentioned that if above actions are ethically or legally questionable, we just replace self-reuse by self-plagiarism.

But what is “ethically or legally questionable”, author did not give us a very clear definition. So we don't have any clear definition of self-plagiarism, either.

----Which is the first shortcoming of this article

After reading some other materials, I got another definition of Self-plagiarism, which is using ones own previously published materials to create a new published material, but not crediting the previous paper as a source.

Current policy of self-plagiarism

The ACM policy : at least 25% of the paper is material not previously published.

The IEEE policy expressly states : plagiarism, self-plagiarism, fabrication, and falsification are “unacceptable”.

How common self-plagiarism is in reality?

Author conducted an experiment (SPlaT) and found that technical reports published in conferences, conference articles recast as journal papers and some similar articles didn't reference with each other.

But there are two problems in this experiment:

First, the publications examined by author are not all papers appear on authors' web sites, and not all papers can be successfully parsed.

Furthermore, when author analyzed the results, he did not give us any data and statistic analysis, just roughly describe the results or list some description.

---- which is the second shortcoming of this article

An information survey

He sent out 30 questionnaires and received 10 responses, and found that most of them encountered cases of self-plagiarism and also reuse their own materials from previously published papers. But they have different ideas on what kind of self-reuse is reasonable and what kind is self-plagiarism.

Question:

Self-plagiarism is very common and has a lot of bad effects on scientific progress.

So what should we do to combat self-plagiarism?

1. Let people realize self-plagiarism's bad effects and hold themselves high standard.
2. More severe policies to supervise.
3. Using self-plagiarism detection tool to detect self-plagiarism.