Self Plagiarism in Computer Science

by
Christian Collberg and Stephen Kobourov

presented by
Jens Trotzky

Summary

- This paper discusses the legitimacy of reused work in publications from different perspectives.

- It tries to create a definition of self-plagiarism.

- It illustrates the characteristics of a publication to be considered a negative example.

Comments

- The paper gives a terminology for self-plagiarism. This supports arguing about types of reused work and makes it easier to categorize published work in terms of legitimacy.

- The paper doesn't offer any statistical data to back up arguments given in the paper.

Comments

- The author gives a couple of examples that makes clear that forming an opinion about reuse of work is not an easy task. He gives a list of responses, though not representative, that shows how different people can argue on this topic.

Definition of Self-plagiarism

- The terms used in the paper are
    - Textual reuse (text/images)
    - Semantic reuse (ideas)
    - Blatant reuse (indistinguishable)
    - Selective reuse (parts)
    - Incidental reuse (not related)
    - Reuse by cryptomnesia (unawareness)
    - Opaque reuse (non-acknowledging)
    - Advocacy reuse (for different community)

Definition of Self-plagiarism

- All these terms help to describe the type of reused work and help to start forming an opinion on whether that kind of reuse clashes with our ethics.

- Some of the terms are intersecting and some are biased towards a certain opinion.

Question

Where does legitimate reuse of own work end and non-ethical self-plagiarism start?

1. Selective reuse in a second paper
2. Opaque Advocacy reuse
3. Technical paper and publication

References

I.   **"Self Plagiarism in Computer Science"** *by* Christian Collberg and Stephen Kobourov, Communications of the ACM, April 2005/Vol. 48, No. 4

II.   **"A Study of Self-Plagiarism in Computer Science"** *by* Christian Collberg, Stephen Kobourov, Joshua Louie and Thomas Slattery (Department of Computer Science, University of Arizona, Tucson, AZ 85721), University of Arizona TR03-02

III. **"ACM Policy on Prior Publication and Simultaneous Submissions"**, http://www.acm.org/pubs/sim_submissions.html

References

IV. **"TurnITin - U.S. Legal Document",**
http://www.turnitin.com/static/legal/legal_document.html

V. **"Self-Plagiarism Tool"**, http://splat.cs.arizona.edu

VI. **"Policy on Self-Plagiarism",** IEEE,
http://www.comsoc.org/dl/net/Self_Plagiarism.pdf

VII. **"EVE2 – Essay Verfication Engine",**
http://www.canexus.com/eve

VIII. **"My DropBox.com",** http://www.mydropbox.com

Any questions?

# Additional slides for questions, etc.

Reality and guidelines

- Some information about self-plagiarism in sponsored projects is available but doesn't give an estimate on overall numbers. [II]

- ACM: "…at least 25% of the paper is material not previously published…" *but* "…is left up to each publication to interpret…" [III]

- IEEE: "unacceptable." [VI]

Types of Reuse

Textual

By Cryptomnesia

Incidental

Semantic

Selective

Opaque

Blatant

Advocacy
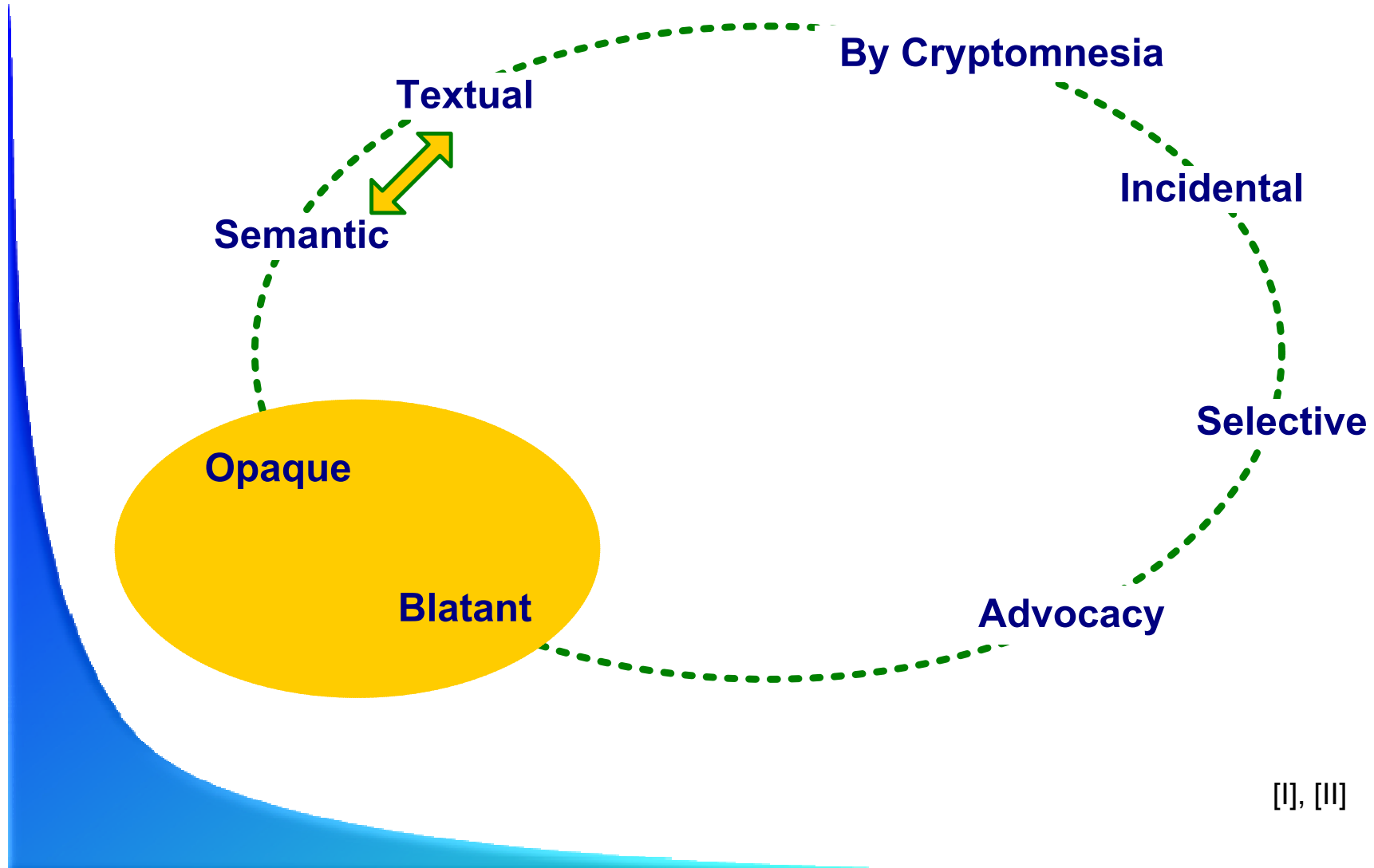
Ethics

- Related to the question of increasing profit for oneself.

- Not a legal question but a question of legitimacy.

- Opinions on this topic vary and are often case based.

Self-Plagiarism

- Majority of people would agree that self plagiarism starts if people reuse their work in a blatant or opaque way.

- The question whether self-plagiarism is a bad thing is answered differently and depending on the type of self-plagiarism.
  - Intentions of the author.
  - Added value and reputation gained.

Tools for Detection

- SPlaT (Self Plagiarism Tool)
  - Uses several resources such as ACM database, Web crawler, etc. to compare a text with existing publications.

- EVE2, TurnItIn, MyDropBox.com
  - Mainly for plagiarism detection based on same techniques.
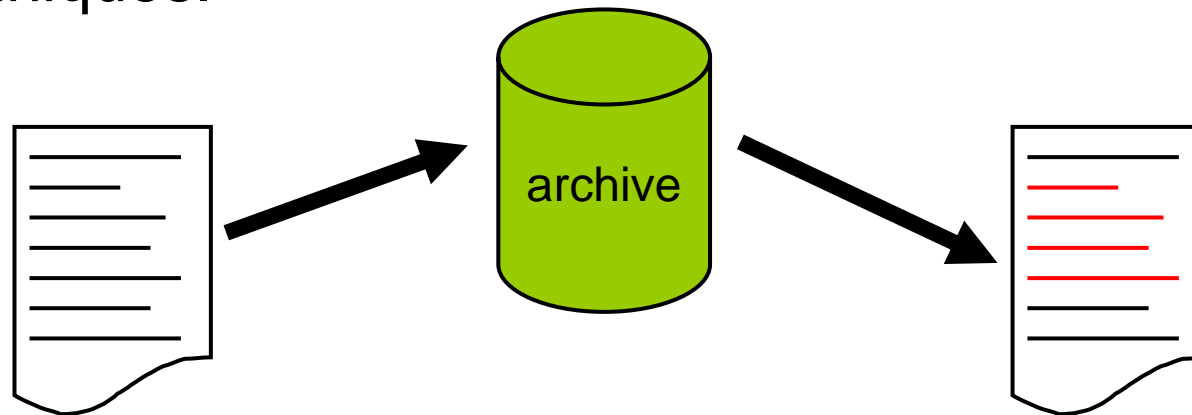
archive

Problems

- Copyright issues for detection tools
  - Archiving articles and using them for comparison.
- Detection of Textual reuse possible.
- Detection of Semantic reuse hard.
- No rules on consequences.

Critique

- No data on number of cases.

- Pointing out importance.

- Difficult point of view on Plagiarism vs. Opaque Self-Plagiarism.

- There is a second paper which is almost identical