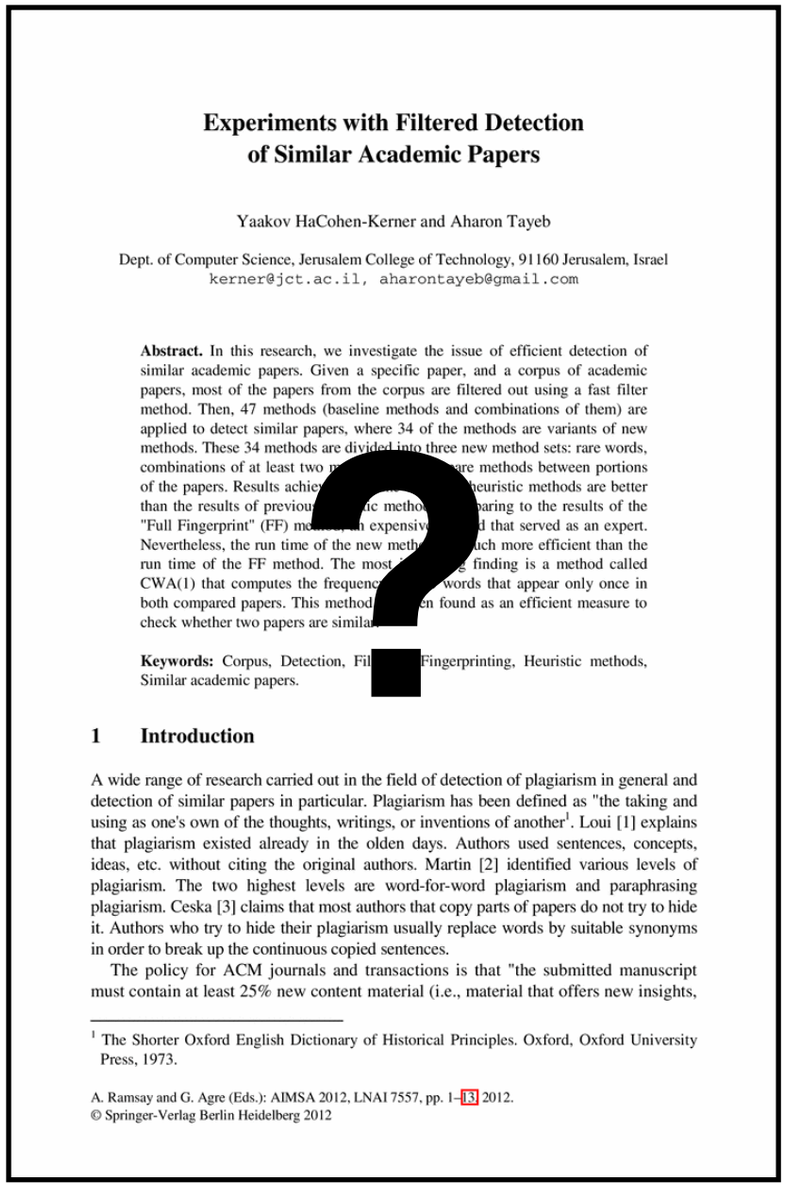
Theoretical Aspects of Modern Security & Privacy Research

Instructor: Prof. Dr. Michael Backes
Advisors: Dr. Robert Künnemann, Kathrin Grosse,
Jonas Schneider

CISPA
Center for IT-Security, Privacy
and Accountability

UNIVERSITÄT
DES
SAARLANDES

How Do You Think Papers Get Published?



Experiments with Filtered Detection of Similar Academic Papers

Yaakov HaCohen-Kerner and Aharon Tayeb

Dept. of Computer Science, Jerusalem College of Technology, 91160 Jerusalem, Israel
kerner@jct.ac.il, aharontayeb@gmail.com

Abstract. In this research, we investigate the issue of efficient detection of similar academic papers. Given a specific paper, and a corpus of academic papers, most of the papers from the corpus are filtered out using a fast filter method. Then, 47 methods (baseline methods and combinations of them) are applied to detect similar papers, where 34 of the methods are variants of new methods. These 34 methods are divided into three new method sets: rare words, combinations of at least two m[illegible]are methods between portions of the papers. Results achie[illegible]heuristic methods are better than the results of previou[illegible]ic metho[illegible]paring to the results of the "Full Fingerprint" (FF) me[illegible]n expensiv[illegible]d that served as an expert. Nevertheless, the run time of the new meth[illegible]uch more efficient than the run time of the FF method. The most [illegible]e finding is a method called CWA(1) that computes the frequenc[illegible]words that appear only once in both compared papers. This method[illegible]en found as an efficient measure to check whether two papers are simila[illegible]

Keywords: Corpus, Detection, Fil[illegible]Fingerprinting, Heuristic methods, Similar academic papers.

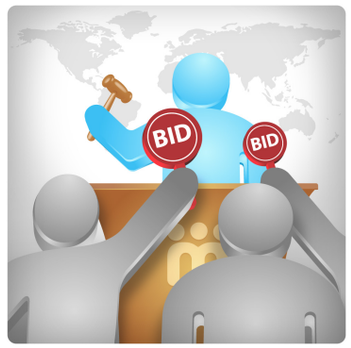
1 Introduction

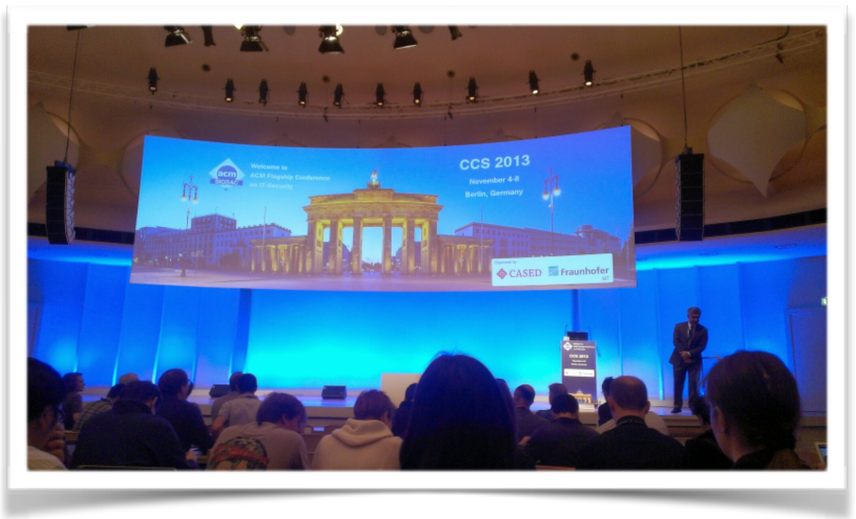
A wide range of research carried out in the field of detection of plagiarism in general and detection of similar papers in particular. Plagiarism has been defined as "the taking and using as one's own of the thoughts, writings, or inventions of another[1]. Loui [1] explains that plagiarism existed already in the olden days. Authors used sentences, concepts, ideas, etc. without citing the original authors. Martin [2] identified various levels of plagiarism. The two highest levels are word-for-word plagiarism and paraphrasing plagiarism. Ceska [3] claims that most authors that copy parts of papers do not try to hide it. Authors who try to hide their plagiarism usually replace words by suitable synonyms in order to break up the continuous copied sentences.

The policy for ACM journals and transactions is that "the submitted manuscript must contain at least 25% new content material (i.e., material that offers new insights,

[1] The Shorter Oxford English Dictionary of Historical Principles. Oxford, Oxford University Press, 1973.

A. Ramsay and G. Agre (Eds.): AIMSA 2012, LNAI 7557, pp. 1–13, 2012.





The Seminar

Simulating a computer-science conference

1. Write and submit a paper
2. Bid on papers and assign papers to reviewers
3. Review papers carefully
4. Write a rebuttal for your submission
5. Meet to discuss submitted papers
6. Prepare the camera-ready version of accepted papers
7. Present accepted papers at the conference

The Seminar

Simulating a computer-science conference

1. ~~Write and submit a paper~~
2. Bid on papers and assign papers to reviewers
3. Review papers carefully
4. Write a rebuttal for your submission
5. ~~Meet to discuss submitted papers~~
6. ~~Prepare the camera-ready version of accepted papers~~
7. Present accepted papers at the conference

Seminar Schedule

- **Kick-off meeting (today)**
  - Bid on papers, start reading assigned papers + submissions
- **Review submission deadlines**
  - Optional submission of first review: **January 16th**
  - All reviews due on **January 23rd**
- **Rebuttal due on January 30th**
- **Slide Review**
  - Make an individual appointment with your advisor
  - Latest one week prior to presentation
- **Paper presentation on Thursday 9th of February (8:30am-11:30pm and 12:30pm-15:30pm)**

Your Tasks

- Bid on all papers
- Review 3 papers
- Write a rebuttal for your paper
- Present your paper (20 minutes + 5 minutes Q/A)

Grading

- Your written work (reviews, rebuttal)
- Your presentation
- Your participation in the papers' discussion

Seminar's Topics

- Formal Methods (4 submissions)
  - supervised by Robert Künnemann, please arrange meeting via brief email to robert.kuennemann@uni-saarland.de
- Adversarial Machine Learning (3 submissions)
  - supervised by Kathrin Grosse, best arrange meeting via brief email to kathrin.grosse@cispa.saarland
- Fully Homomorphic Encryption+Secure Multiparty Computation (3 submissions)
  - supervised by Jonas Schneider, come to 3.16 anytime between 10:00 — 17:00

Formal Methods

1. Policy Auditing over Incomplete Logs: Theory, Implementation and Applications *Oloutoyin Salomon Laleye*

2. Causes and Explanations: A Structural-Model Approach — Part I: Causes *Turbat Ganbold*

3. Program Actions as Actual Causes: A Building Block for Accountability *Dhiman Chakraborty*

4. CoSP: A General Framework For Computational Soundness Proofs *Sharmeen Rehan*

Adversarial Machine Learning

6. Cryptography and Machine Learning *Vincent Ogwara*

7. Adversarial Perturbations Against Deep Neural Networks for Malware Classification *Marius Steffens*

8. Transferability in Machine Learning: from Phenomena to Black-Box Attacks using Adversarial Samples *Nadisha-Marie Aliman*

FHE and MPC

11. A Guide to Fully Homomorphic Encryption
    Bakhtiar ali shah

12. Secure Multiparty Computation for Privacy-Preserving Data Mining *Siavash Riahi*

13. Non-Interactive Verifiable Computing: Outsourcing Computation to Untrusted Workers
    Kevin Morio

How to Write a Review

A Guide for New Referees in Theoretical Computer Science*

Ian Parberry†
Department of Computer Sciences
University of North Texas

Goal of the Presentation

- You should convey (to the audience!)
  - goal and applicative context of your paper
  - contributions of the paper
  - scientific context (e.g., related work, prior state of the art)
  - ideally: a balanced assessment beyond “limitations” section
- Food for discussion:
  - prepare at least one question to initiate discussion

Getting Good Grades

- Research literature independently and relate what you find to your paper
- Get help if necessary – not asking for help is foolish, not smart
- Deep understanding of your paper
- Well balanced critical assessment — bashing is much easier than balanced discussion

What to Do Next?

- Read the e-mail for your HotCRP account
- Read further instructions (in a second e-mail) with information about how to provide your preferences in the HotCRP system
- Give your preferences on papers
  - Read the abstracts of all papers
  - Based on the abstract/topics, try to read in more detail, and understand, a subset of interesting papers
- Read your own paper (just chosen by you)