SNITCH
Spotting and Neutralizing Internet Theft by Cheaters
Sebastian Niezgoda and Thomas P. Way

Abstract
Plagiarism of material from the Internet is a widespread and growing problem. Computer science students, and those in other science and engineering courses, can sometimes get away with a cut and paste approach to assembling a paper in part because the expected style of technical writing is less expository than in liberal arts courses. Detection of cut and paste plagiarism is tedious and time-consuming when done by hand, and can be greatly aided by automated software tools. This paper reports on the design of a software tool called SNITCH that implements a plagiarism detection algorithm using the Google Web API. Issues related to plagiarism detection software are discussed and empirical results of a performance and accuracy study are presented.

How does it work?
- Simple interface
  - User adds one or more papers to be analyzed
  - User selects the paper(s) to be analyzed and clicks Analyze
  - Results are presented as web pages

Workflow
- Divide the document into windows
- Find possibly plagiarized passages
- Rank passages
- Search for passages on Google
- Generate results

Definitions
- Window
  - a set of words
  - could be a short sentence or a section of a sentence
- Window size
  - Number of words in a window
  - Window weight
    - Average word length \cdot window size

Window Creation
- How SNITCH does it:
  - Phrase
  - Our research
  - Our research in
  - Our research in computer
  - Our research in computer architecture
  - Our research in computer architecture covers memory hierarchy, branch prediction, superscalar implementation, as well as SMT and multithreaded processors. This survey paper explains and classifies the various multithreading techniques in research and in commercial microprocessors and compares multithreaded processors with chip multiprocessors.

The Algorithm
- Sliding window algorithm
  - Try to mimic how a human reads
  - Starting with the first word and following words to window until window size has been reached; then create a new window starting with the second word; and so on...

Sliding Window Example
- What SNITCH found:
  - Our research in computer architecture covers memory hierarchy, branch prediction, superscalar implementation, as well as SMT and multithreaded processors. This survey paper explains and classifies the various multithreading techniques in research and in commercial microprocessors and compares multithreaded processors with chip multiprocessors.

Google
- Used to search for passages on-line
- Provides an environment for searching directly from Java
- Free to use
- Limited
  - 1000 queries per day

Results & Conclusions
- It works!
- It’s quick!
- Constant improvement battle
  - What is the most accurate and efficient way to select just the passages that are most likely to be plagiarized?