



# PLAGIARISM DETECTION IN HASKELL ASSIGNMENTS

---

CAT HUSTON



# INTRODUCTION

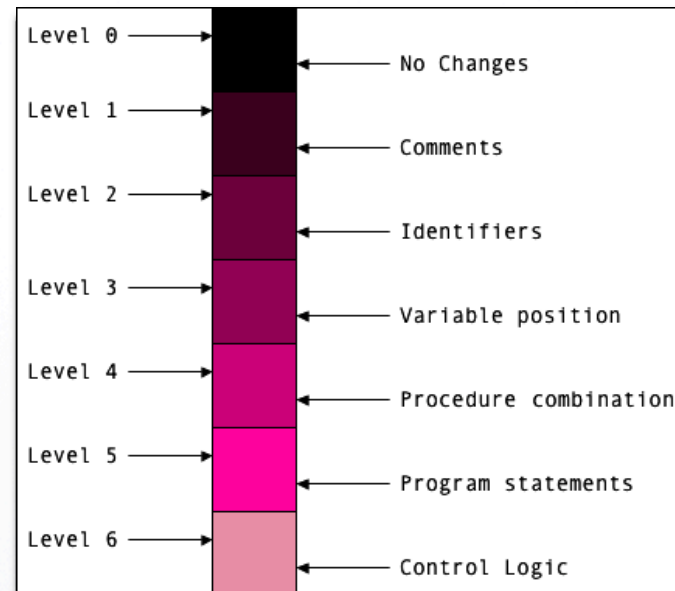
- Plagiarism: What and Why?
- Reducing Plagiarism
- Detecting Plagiarism
- Differences between Java and Haskell
- Proposed solutions
- Conclusions



# PLAGIARISM: WHAT AND WHY? PART 1

To take (the work or idea of someone else) and pass it off as one's own.

Compact Oxford English  
Dictionary



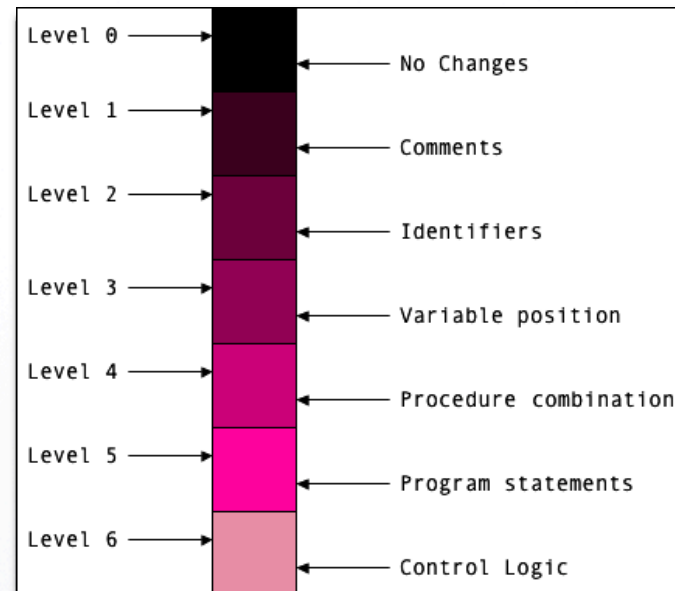
Plagiarism Scale:  
Adapted from  
Hamblen and Parker,  
1998.



# PLAGIARISM: WHAT AND WHY? PART 1

“Plagiarism is the practice of (dishonestly) claiming or implying original authorship of material which one has not actually created, such as when a person incorporates material from someone else's work into his own work without attributing it. Within academia, plagiarism is seen as academic dishonesty and is a serious and punishable academic offense.”

Wikipedia



Plagiarism Scale:  
Adapted from  
Hamblen and Parker,  
1998.



# PLAGIARISM: WHAT AND WHY?

## PART 2

- Course Organization
- School Environment, Curriculum and Computer Consultants
- Hardware Environment
- Student Related Factors

Wegner, “Factors Contributing to Copying”



# REDUCING PLAGIARISM

- Minimize the causes of plagiarism:
  - Improve course organization
  - Maximize hardware availability
  - Provide additional help to struggling students
- Detect instances of plagiarism.



# DETECTING PLAGIARISM

- Google's top 5 (non-sponsored) hits for "plagiarism detection" programming
  - Glatt Plagiarism Screening Program: for written assignments
  - The Plagiarism Resource Site: for written assignments
  - Moss: for use in comparing the similarity of programs
  - Discussion of plagiarism detection in essays and its implications
  - News article discussing the threat of detection as a deterrent to plagiarism
- So far, plagiarism detection has been focussed on the plagiarizing of written assignments, rather than code.

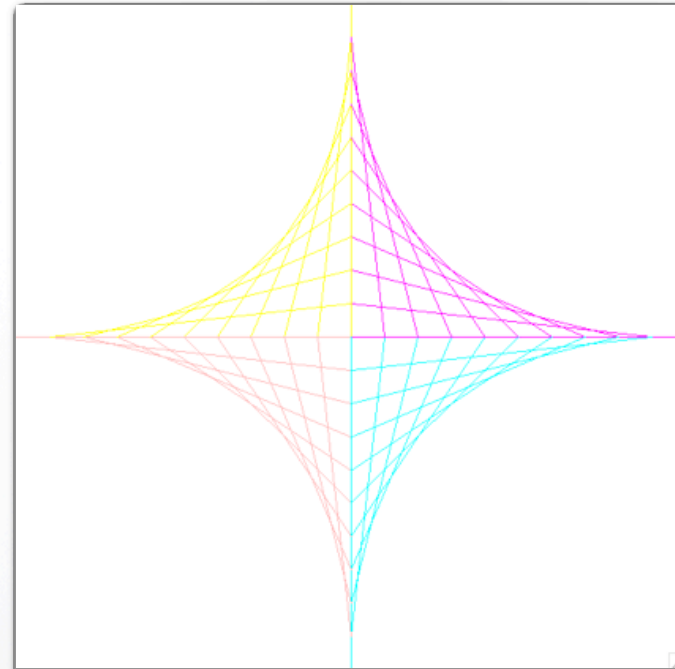


# THE DIFFERENCE BETWEEN JAVA AND HASKELL:

First Java assignment for  
2005-2006.

Graphics

Question 1



1. Anti-clockwise  
axes

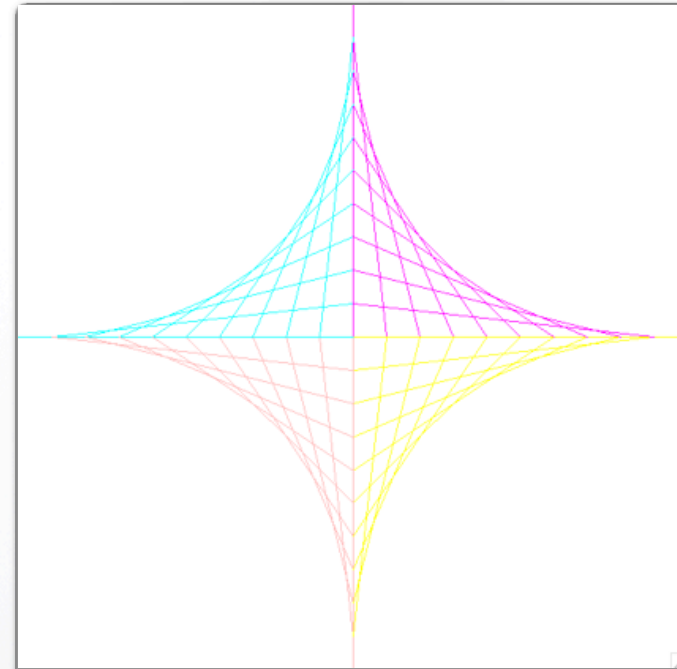


# THE DIFFERENCE BETWEEN JAVA AND HASKELL:

First Java assignment for  
2005-2006.

Graphics

Question 1



2. Clockwise axes

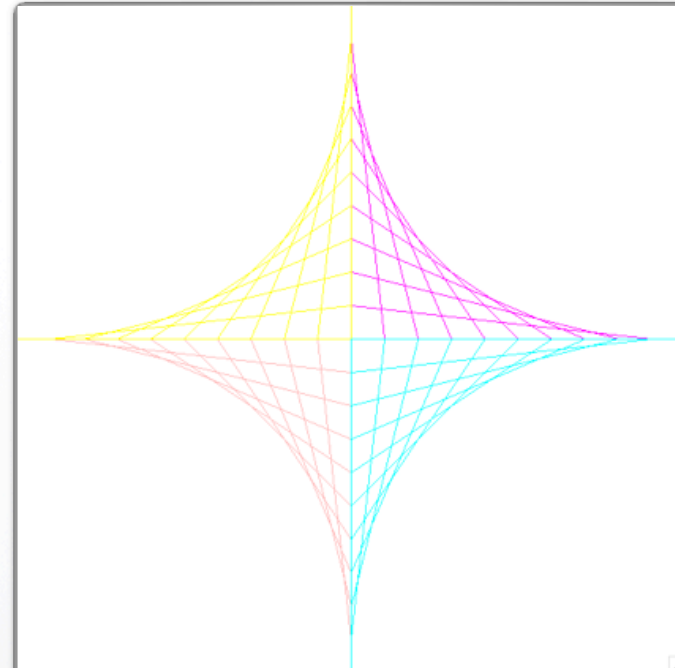


# THE DIFFERENCE BETWEEN JAVA AND HASKELL:

First Java assignment for  
2005-2006.

Graphics

Question 1



3. Emphasis 1

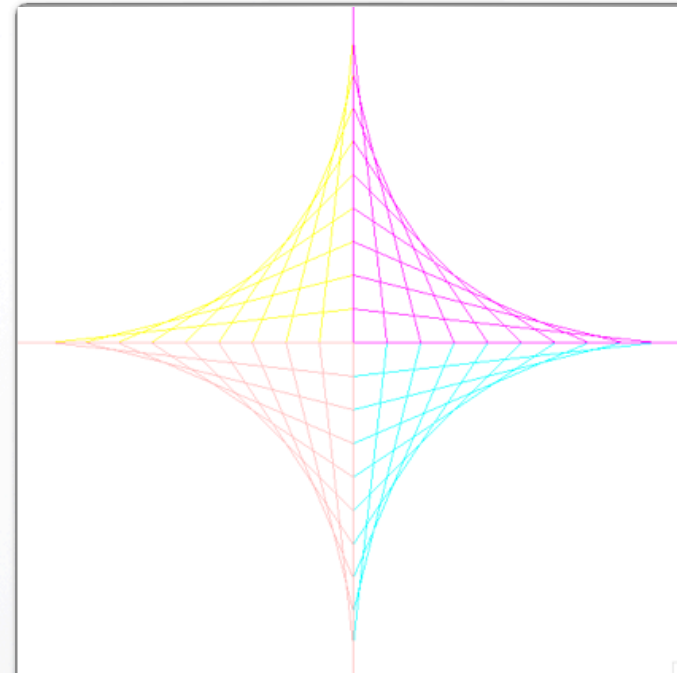


# THE DIFFERENCE BETWEEN JAVA AND HASKELL:

First Java assignment for  
2005-2006.

Graphics

Question 1



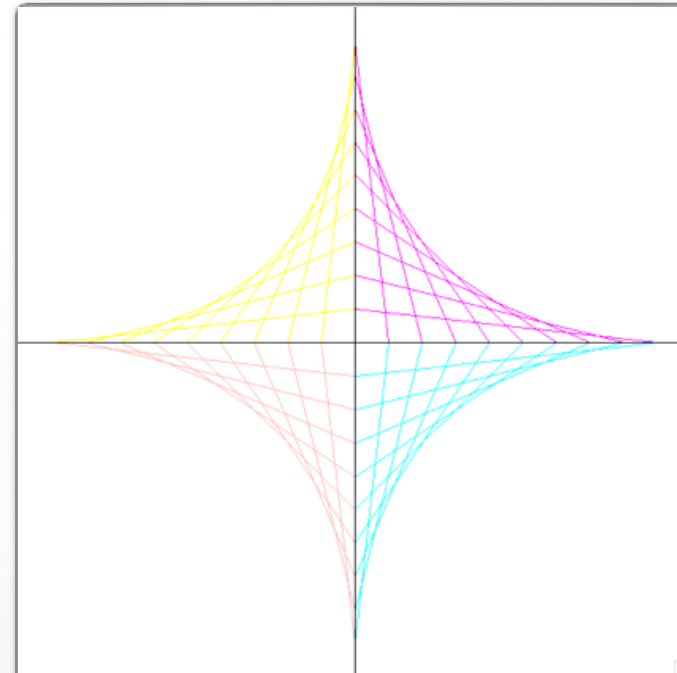
4. Emphasis 2



# THE DIFFERENCE BETWEEN JAVA AND HASKELL:

First Java assignment for  
2005-2006.

Question 1



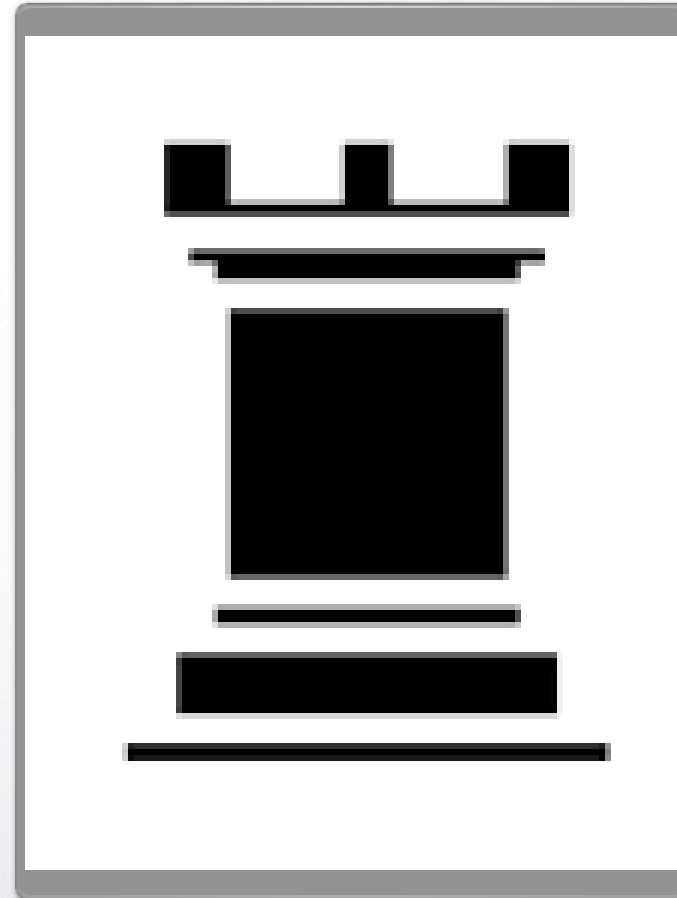
5. Axis drawn  
separately



# THE DIFFERENCE BETWEEN JAVA AND HASKELL:

First Haskell assignment for  
2005-2006.

`blackRook = invert rook`





# PROPOSED SOLUTIONS

- MOSS
- Syntactic Evaluation
- Compare test output
- Apply transformations



# MOSS

- Measure of Software Similarity
- Works by taking a “document fingerprint”
  - n-grams
  - ignores variable names
- Limited effectiveness for the first year Haskell assignments
  - Short answers
  - Structure of the assignment means that two students can produce very similar assignments independently.



# SYNTACTIC EVALUATION

- For instance,
  - Identical variable names
  - Redundant brackets
- Problems:
  - Would only detect plagiarism at levels 0 and 1
  - Sometimes we expect variable names to be identical
  - Determining when similarity is significant automatically would be non-trivial

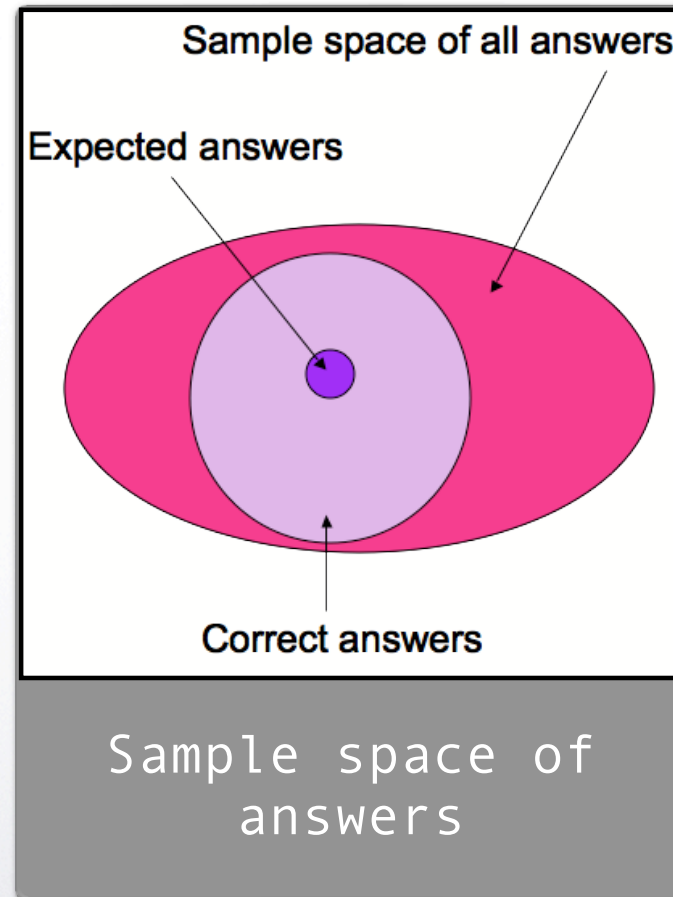


# TRANSITIONS

Determine by transitioning on programs where they lie in the sample space.

The further a program lies from the expected answer the more suspicious any similarity is.

This is a difficult problem to solve, it is likely that any solution would take too long to compute.





# CONCLUSIONS

- Plagiarism detection is different in Haskell than in Java.
  - Aspects of the language
  - Aspects of the course
- There are various solutions to this
  - One potential solution is looking at patterns in test output
  - Copied solutions may have similar behaviour



# SOURCES

- “Plagiarism Detection”, David Daniel Madden
- Wikipedia, [en.wikipedia.org/wiki/Plagiarism](http://en.wikipedia.org/wiki/Plagiarism)
- Compact Oxford Dictionary, [www.askoxford.com](http://www.askoxford.com)
- Glatt Plagiarism Screening Program  
[www.plagiarism.com/screening.htm](http://www.plagiarism.com/screening.htm)
- The Plagiarism Resource Site,  
[plagiarism.phys.virginia.edu](http://plagiarism.phys.virginia.edu)
- Moss, [theory.stanford.edu/~aiken/moss](http://theory.stanford.edu/~aiken/moss)
- Plagiarism-Detection Software Stems Students' Use Of  
"Paper Mills", [www.sciencedaily.com/releases/  
2002/05/020501073710.htm](http://www.sciencedaily.com/releases/2002/05/020501073710.htm)
- QuickCheck [www.md.chalmers.se/~rjmh/QuickCheck/](http://www.md.chalmers.se/~rjmh/QuickCheck/)