

# In-Class Activity 2: Web Crawling

# Overview

- Use a common web testing technique
- How to systematically explore web pages?
- Solution: Crawl a web application
  1. Use Crawljax ([www.crawljax.com](http://www.crawljax.com))
  2. Download via Maven

# Goals for IAC 2

1. All goals from IAC 1
2. Perform a complete crawl of a web app
3. Verify/report properties on each page

Know how to do these things!

Blank

# Implement the Following

1. Automated crawl of bookstore
2. Achieve 50% statement coverage (10 min max)
  1. Exclude Common.jsp and your Cobertura servlet from the report
3. Produce a report that contains a line of the form `<Servlet Name> : <tag>` , where `<tag>` is every `<form>`, `<img>`, or `<a>` tag found in the app.
4. Report count of tags of each type found per servlet.

# Grading

100 total points

- 15 points for the quiz
- 20 points for in-class checkoff (late credit until 9/30)
- 15 points for complete automation
- 40 points for coverage (partial credit ok, but must have minimum of 40% coverage)
- 10 points for report generation

Bonus, 10 points each:

- Highest coverage in-class
- Highest coverage by Monday (9/30)

# Rules

- Must be a real crawl
- 10-20mins (strict)
- General input rules ok, but not ok to presuppose knowledge of each page

# Submit

1. Pom file
2. Shell script
3. Java files
4. Coverage report
5. Tag report