Recovering management information from source code

Peter Kampstra
Łukasz Kwiatkowski
Vrije Universiteit Amsterdam, De Boelelaan 1081a,
1081 HV Amsterdam, The Netherlands
{pkampst,lukasz}@few.vu.nl

http://www.cs.vu.nl/equity
Contents

• EQUITY project
• Motivation
• Proposed areas of analyses
• Summary & discussion
EQUITY

• How can we use existing data to identify and maximize IT Profits?

• IEEE Equity
  (http://www.cs.vu.nl/equity2007)

• X Verhoef
EQUITY (2)

• Source code one of the inputs

• >50 million lines of Cobol

• Project management data

• Since Sep 2005, ongoing research!
Objectives

• Define source code metrics useful for management
• Couple metrics with management information
• Improve cost prediction models accuracy
IT portfolio size

• How much IT do we have?
• Larger systems are more complex
• Used for cost estimation
Comparing portfolios

Portfolio I

Portfolio II

occurrences

occurrences

10 May 2006  
Exploring QUantifiable Information Technology Yields  
Recovering management information from source code
Source code volatility

• Potential causes:
  – Error-prone modules
  – Implementing new features
  – Quality improvements (reengineering)

• Extract version data from comments
Portfolio volatility

High volatility \(\rightarrow\) high costs?

![Bar chart showing average number of module changes across information systems.](chart.png)
Complexity + volatility

- Indication of error-prone modules
- Input for data-mining
Portfolio age

- Maintainability

- Less knowledge is available for older modules
Dead code

- S Klusener’s work
- Unreachable code should not be paid for

% of dead files
% of dead SLOC

cluster

% of dead code

0 10 20 30 40 50 60 70

1 2 3 4 5 6 7 8
Points needing change

• Experience from 9210 project:

• Expected cost of change reduced by a factor 4
Other aspects

• Source code quality
• Architectural quality
Definition

Outsourcing = Leaving torn and dirty underwear with a laundry and expecting new clothing in return

(P. Strassmann)
Summary

• Much data hidden inside source code
• Management information can be extracted

• Searching for data...
Discussion

• Which problems are there when extracting data?
• Which metrics and graphs are interesting?
• ....