Empirical Methods Overview

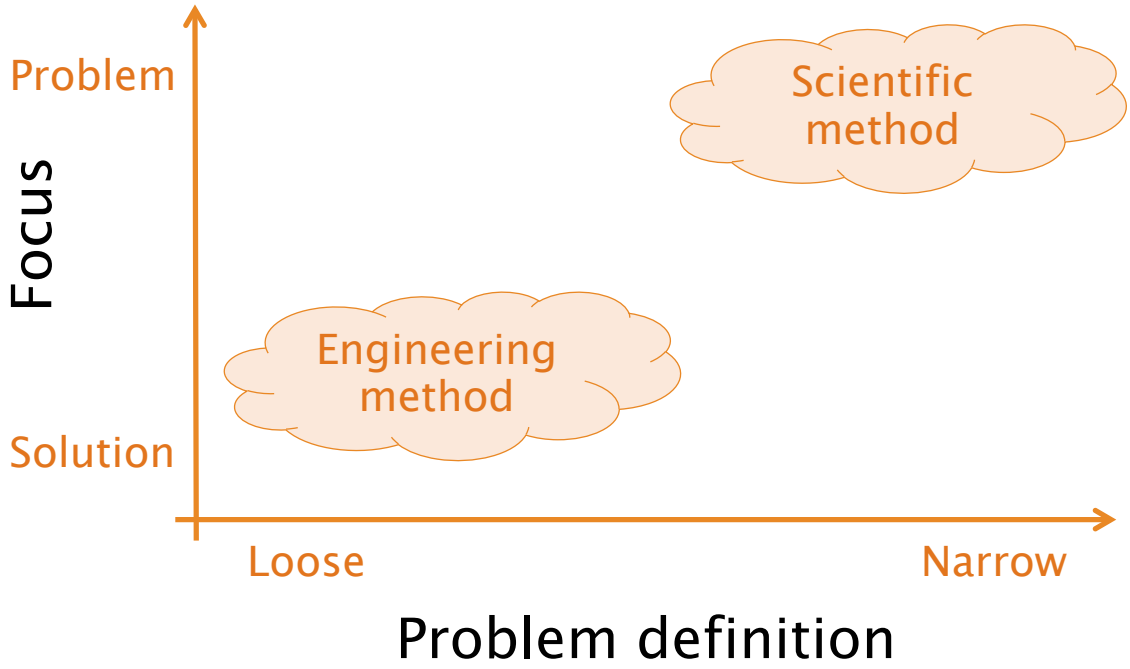
(01OPJIU) Empirical Methods in Software Engineering

http://softeng.polito.it/EMSE/

SoftEng
http://softeng.polito.it

Version 1.2.2

Research methods



Engineering method

The use of heuristics to cause the best change in a poorly understood situation within the available resources

Koen, 2003

Heuristic

To find or to discover (Εὑρίσκω)

on the basis of

Experience (Ἐμπειρία)

- How do we build that experience?
  - Casual and Anecdotal
  - Systematic and Scientific

Quality Improvement Paradigm

Experience Factory

Quality improvement

- Any successful business requires a combination of technical and managerial solutions

Software Discipline Premises

- Evolutionary and experimental
- Development as opposed to production
- Makes use of technologies that are ultimately human based

Software is Software is Software?

- No!
- All software is not the same
  - Process is a variable
  - Goals are variable
  - Content varies
  - …

Context does matter

- "Technology is not universally good or universally bad, only more appropriate in some circumstances and for some organizations."

[T.Dyba, B. A. Kitchenham and M. Jorgensen, "Evidence-based software engineering for practitioners,"]

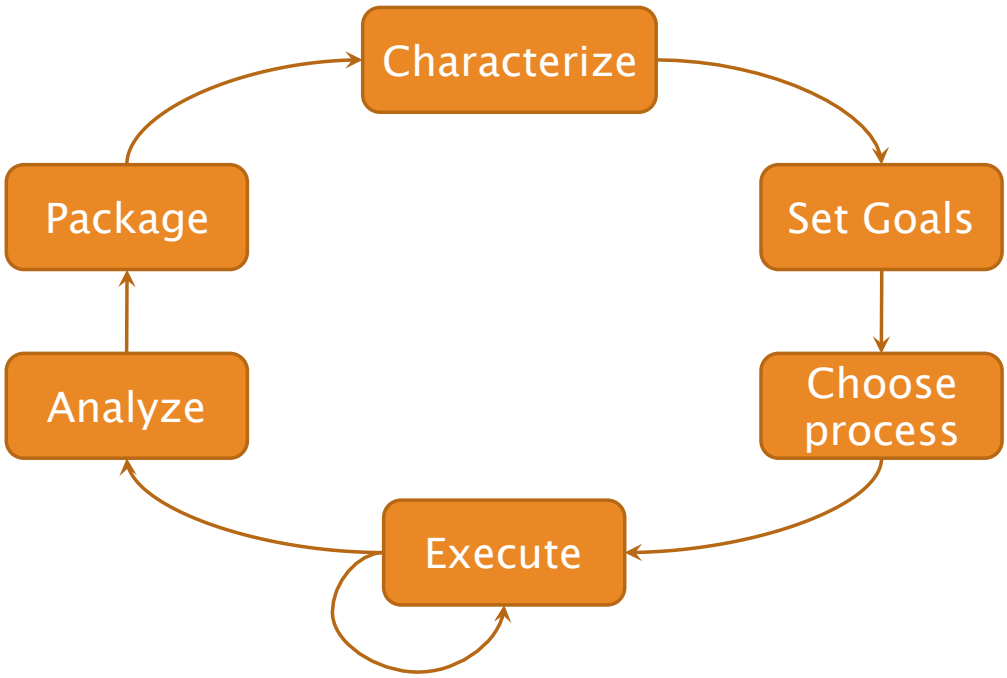
Models…

- In general there's a lack of models that allow us to reason about
  - Process
  - Product
- The few available are not always understood in context

Experience Factory/QIP

- Quality improvement in the context of software business
- Mechanism for continuous improvement
  - Experimentation
  - Packaging
  - Reuse
- Combine scientific and engineering methods

QIP Steps



QIP

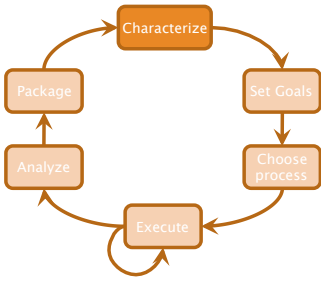
- Characterize
  - The current project and its environment w.r.t. known models and metrics
- Set the quantifiable goals
  - for project success and improvement
- Chose
  - Appropriate process model, methods, and tools, and customize them
- Execute
  - And collect and validate prescribed data

QIP

- Analyze the data
  - Evaluate current practices, determine problems
  - Record findings, make recommendations for future
- Package experience
  - Updated and refined model
  - Other forms of structured knowledge

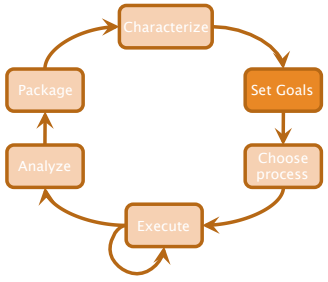
Characterize Project and Environment

- Factors
  - People
    - team size, expertise, organization, experience
  - Problem
    - Application domain, change proneness
  - Process
    - Life cycle, methods, notations
  - Product
    - Deliverables, size, qualities
  - Resource
    - Target platforms, calendar, budget



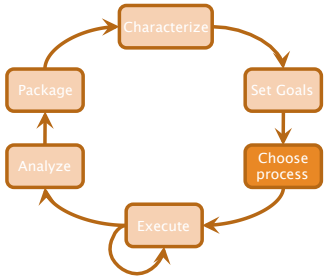
Define goals

- Goals
  - Both product and process
  - Measurable
  - Driven by models
- Techniques
  - Goal Question Metric (GQM)
  - Quality Function Deployment (QFD)



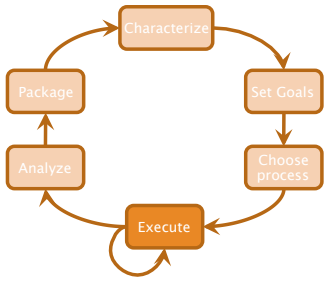
Chose execution model

- Most suitable for
  - Context
  - Environment
  - Characteristics
  - Goals
- Based on previous knowledge
- Tailoring is typically required



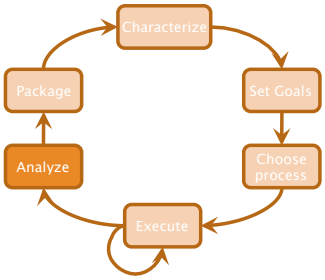
Execute process

- Reuse of packaged experiences
- Data collection
  - Integrated into the process
  - Automatic
  - Manual



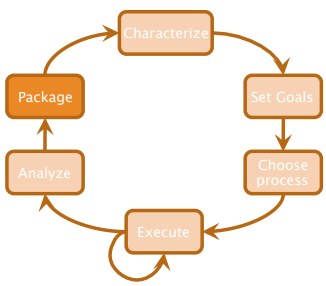
Analyze data

- Characterize and understand
  - E.g. Which is the typical source of errors?
- Evaluate
  - E.g. What is the test plan coverage?
- Predict and control
  - E.g. What is expected cost?
- Motivate and improve



Package results

- Define and refine
  - Models
    - Mathematical models
    - Informal relationships
    - Algorithms
  - Baselines
- Packaging means also
  - Training
  - Deployment
  - Institutionalization



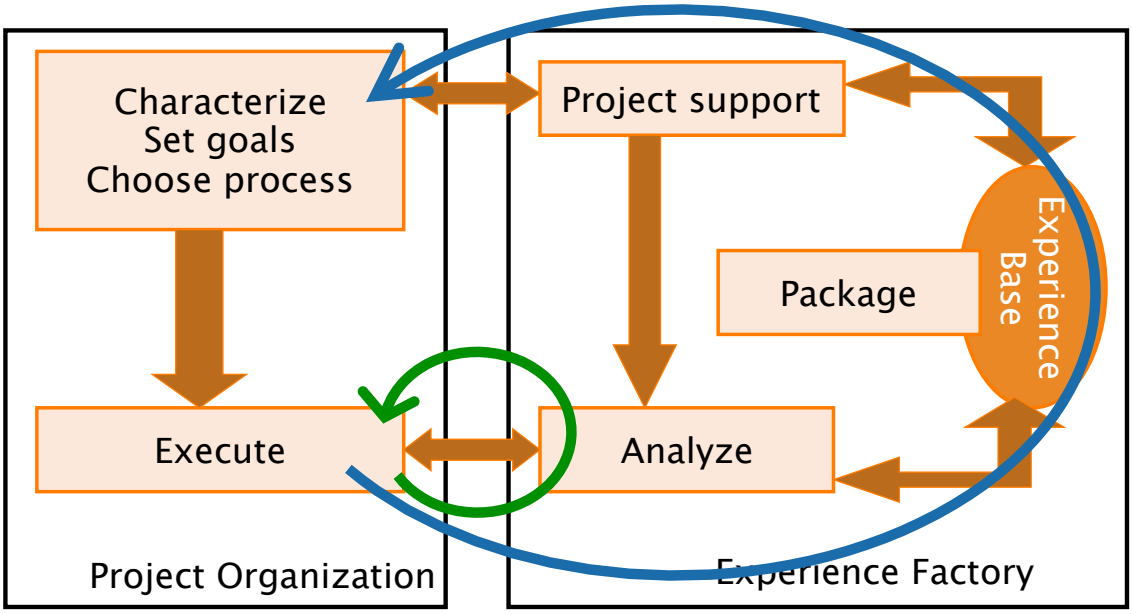
Experience Factory

- Actuating QIP requires an adequate company organization
- Key elements
  - Learning
  - Experience Models
  - Experience Base
  - Reuse

Experience Factory

- Two different focuses
  - Software development
  - Systematic learning
- Organization made up of
  - Project organization
  - Experience factory

Experience Factory

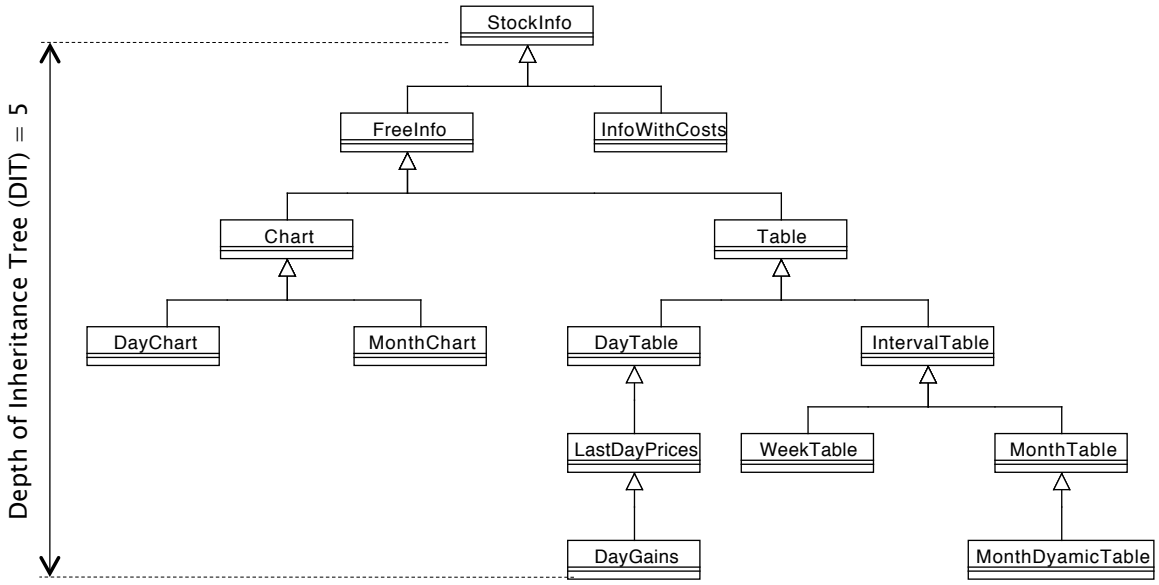


EVIDENCE BASED SOFTWARE ENGINEERING

EBSE

- Practitioners can have difficulty making informed decisions
- Often adopt a technology without evidence or ignore others
  - Enthusiasts of object-oriented programming were initially keen to promote the value of hierarchical models.
  - Only later did experimental evidence reveal that deep hierarchies are more error prone than shallow ones

Ex. Inheritance Hierarchies



Adapted from Prechelt et al. 2003

EBSE Steps

1. Convert a relevant problem or information need into an answerable question.
2. Search the literature for the best available evidence to answer the question.
3. Critically appraise the evidence for its validity, impact, and applicability.
4. Integrate the appraised evidence with practical experience and the customer's values and circumstances to make decisions about practice.
5. Evaluate performance and seek ways to improve it.

Ask an answerable question

- How to appraise and apply methods, tools, and techniques in practice
- Three components:
  - The main intervention or action you're interested in
  - The context or specific situations of interest
  - The main outcomes or effects of interest

Find the best evidence

- Separate
  - Question you want to answer
  - Question implemented in the search terms
  - Questions answered in the studies found
- Different sources
  - customers or the software's users,
  - colleagues or experts,
  - learned as student or in professional courses
  - search for re- search-based evidence

Critically appraise the evidence

- Is there any vested interest?
- Is the evidence valid?
- Is the evidence important?
- Can the evidence be used in practice?
- Is the evidence consistent with the evidence in other available studies?

Apply the evidence

- Integrate the evidence with
  - your practical experience
  - customers' requirements
  - knowledge of the concrete situation's specific circumstances,
- Employ the evidence in your decision making

Evaluate performance

- Assess whether process change has been effective
  - task with an identifiable purpose
  - some metrics to measure performance
- After-action review meetings (AAR)
  - What was supposed to happen?
  - What actually happened?
    Why were there differences?
  - What did we learn?

EMPIRICAL METHODS

Empirical method

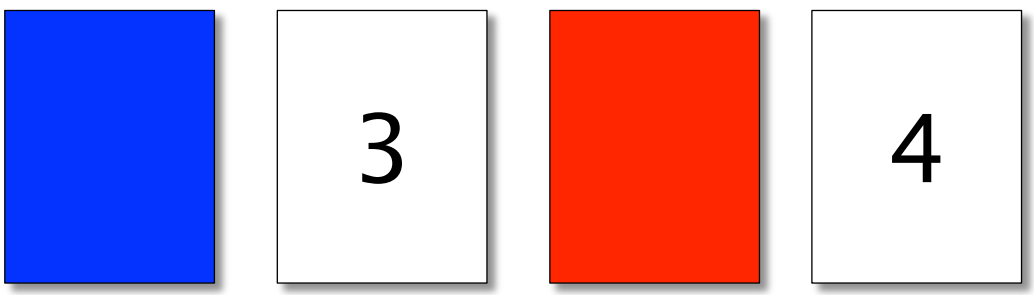
- Gather observations
- Form a conjecture
- Deduce a prediction from that conjecture
- Test by experiment

Testing predictions

- A quick experiment…

Conjecture

- If a card shows an even number on one face, then its opposite face is red
  - Which card(s) you need to turn over to verify the above conjecture holds true?



Conjecture

- If you are drinking alcohol then you must be over 18
  - Which card(s) you need to turn to verify the above conjecture holds true?

A

B

C

D

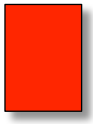
Image by vectorix / stockarch.com

Testing

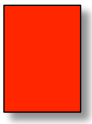
- Conjecture: if even then red
- Testing
  - A: If  then 

  - B: If 

then 

  - C: If  then 

  - D: If 

then 

Wason Selection Task

- Wason selection task
  - Devised in 1966
- Expose a paradox of material implication
  - Implication in natural language
  - Formal implication in logic

Argumentation

- *Modus ponens* (implication elimination)
  - If P, then Q
  - P
  - Therefore Q
- *Modus tollens* (denying the consequent)
  - If P, then Q
  - Not Q
  - Therefore, not P

Formal fallacies

- Affirming the consequent
  - If P, then Q
  - Q
  - ~~Then P~~
- Denying the antecedent
  - If P, then Q
  - Not P
  - ~~Then Not Q~~

Hypothetic-deductive model

- Gather observations:
  - Consider the problem and try to make sense of it. Look for previous explanations.
- Form a conjecture
  - When nothing else is yet known, try to state an explanation.
- Deduce a prediction from that conjecture
  - Assume conjecture is true, what consequences follow?
- Test by experiment
  - Look for the opposite of each consequence.

Falsification

- I believe that we do not know anything for certain, but everything probably.

  (Christiaan Huygens)

- No amount of experimentation can ever prove me right; a single experiment can prove me wrong.

  (Albert Einstein)

- In so far as a scientific statement speaks about reality, it must be falsifiable: and in so far as it is not falsifiable, it does not speak about reality.

  (Karl Popper)

Falsification and confirmation

- A test that could and does run contrary to predictions of the hypothesis is taken as a falsification of the hypothesis.

- A test that could, but does not run contrary to the hypothesis corroborates the theory.

Scientific method

- Define a question
- Gather information and resources (observe)
- Form an explanatory hypothesis
- Test the hypothesis by performing an experiment and collecting data in a reproducible manner
- Analyze the data
- Interpret the data and draw conclusions that serve as a starting point for new hypothesis
- Publish results
- Retest (frequently done by other scientists)

TYPE OF STUDIES

Qualitative vs. Quantitative

- Qualitative studies
- Quantitative studies

Qualitative vs. Quantitative

Feature	Qualitative	Quantitative
Goal	Provide knowledge of an organization and/or a problem and its solution	Generalize the results of a case study to the population of interest
Motivation	Exploratory	Confirmatory
Size	Small	Large
Data collection	Unstructured data	Structured data
Data analysis	Non statistical	Statistical

Qualitative research

- Ethnographic studies
- Grounded theories
- Action research
- Case study research

Quantitative research

- Experiment
- Observational
  - Survey
  - Longitudinal studies
- Case studies

Classification dimensions

- Realism
- Time
- Generality
- Multiplicity
- Control
- Randomness

Realism

- In vivo
- In vitro

Time

- Cross-sectional
  - Takes place at a single point in time.
    - Taking a 'slice' or cross-section of the observed phenomenon
- Longitudinal
  - takes place over time
    - Taking at least two (and often more) waves of measurement

Generality

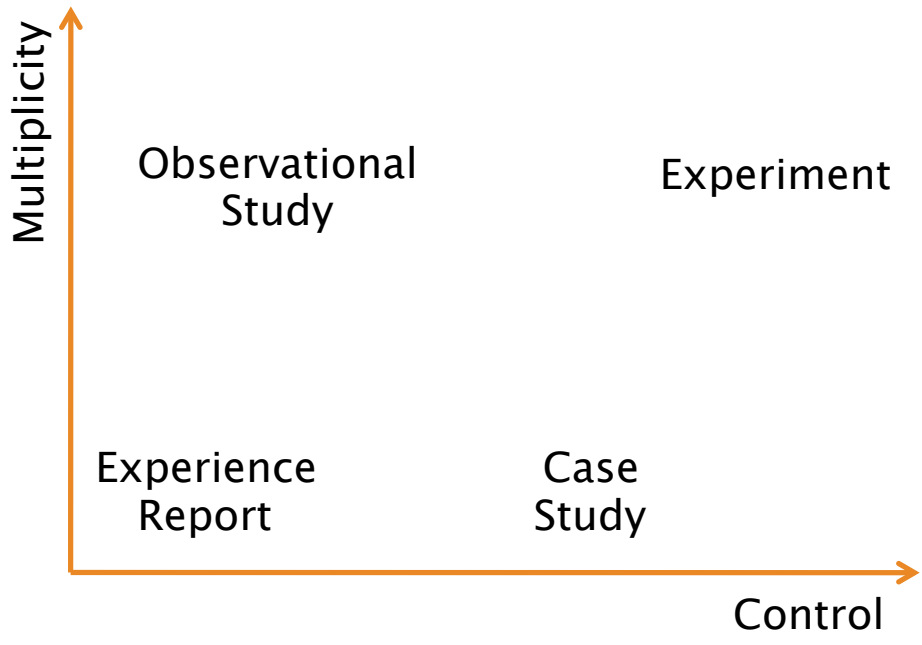
- Nomothetic
  - Derive laws that explain objective phenomena in general
  - Typical for the natural sciences.
- Idiographic
  - Understand the meaning of contingent, unique, and often subjective phenomena.
  - Typical for the humanities.

Multiplicity and control

- Multiplicity:
  - whether several cases are observed.
- Control:
  - whether control has been exercised on the main factor and on the context factors and cofactors.

Tonella et al. 2007

Types of studies



Randomness

- *Randomized experiment:*
  - an experiment in which subjects are assigned to receive the treatment or an alternative condition by a random process.
- *Quasi-Experiment:*
  - an experiment in which subjects are not assigned to conditions randomly, e.g. by convenience sampling or voluntary assignment.

Question type

- Descriptive.
  - Describe what is going on or what exists.
    - E.g. Public opinion polls that report the proportion of people who hold various opinions.
- Relational.
  - Look at the relationships between two or more variables.
    - E.g. How the salary expectations is linked to education years.
- Causal.
  - Determine whether one or more variables causes or affects one or more outcome variables.
    - E.g. the adoption of a technology improves quality

Experiment

- Identifying precise relationships between chosen variables via a designed lab situation, using quantitative analytical techniques, with a view to making generalizable statements
  - The solution and control of a small number of variables which may then be studied intensively.
  - Oversimplification of the experimental situation and the isolation from most of the factors that are found in the real world.

Survey

- Snapshots at a particular point in time from which relationship inferences are made using quantitative analytical techniques
  - Greater number of variables studied; Description of real-world situations; More appropriate generalizations
  - Little insight obtained about causes/ processes behind the phenomena
  - Possible bias in respondents, researcher.
  - Must be rechecked against behavior on live projects.

Case study

- Describe relationships which exist in reality, usually within a single organization. They are helpful to detect, develop, refine frames of reference
  - Captures the local situation in greater detail and with respect to more variables than is possible with surveys
  - Lack of control of variables.
  - Different interpretations by different people.
  - Unintentional biases and omissions in the description

Grounded theory

- Instead of starting with a theory, begins with an area of study and what is relevant to that area is allowed to emerge (Glaser and Strauss 1967)
  - Discovery of unexpected patterns. Utilization of large bodies of situational data not having individual significance.
  - Sensitive to thoroughness and skills of individual researcher

Ethnographic studies

- Gather empirical data on a group of people (sw dev team), aims to describe the nature of those who are studied.
  - Data collection is often done through participant observation, interviews, questionnaires, etc.

Longitudinal studies

- Calls for systematic data collection over a long period of time, to reveal ongoing patterns (or discover new patterns).
  - Can use multiple data collection methods
  - Ability to reexamine data.
  - Richness and amount of data collected

Bibliography

- Koen, Billy Vaughn. (2003). Discussion of The Method: Conducting the Engineer's Approach to Problem Solving. New York: Oxford University Press
- Popper, Karl (1959). The Logic of Scientific Discovery. Routledge
- T. Dyba, B. A. Kitchenham and M. Jorgensen, "Evidence-based software engineering for practitioners," in IEEE Software, vol. 22, no. 1, pp. 58-65, Jan.-Feb. 2005.
- Basili, V. 1995. The Experience Factory and Its Relationship to Other Quality Approaches. In Advances in Computers, vol 41. Academic Press, 65-82

Bibliography

- Potts, C. 1993. Software-engineering research revisited. In IEEE Software, vol. 10(5), September, IEEE Computer Society, 19-28.
- Tonella P., Torchiano, M., Du Bois, B., Systa, T. 2007. Empirical studies in reverse engineering: state of the art and future trends. In Empirical Software Engineering, Vol. 12(5), Springer, 551-571.
- Glaser, B, Strauss, A., *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine de Gruyter, NY, 1967

- Lutz Prechelt, Barbara Unger, Michael Philippsen, Walter Tichy, A controlled experiment on inheritance depth as a cost factor for code maintenance, Journal of Systems and Software, Volume 65, Issue 2, 15 February 2003, Pages 115-126
- R. Harrison, S. Counsell, R. Nithi, Experimental assessment of the effect of inheritance on the maintainability of object-oriented systems, Journal of Systems and Software, Volume 52, Issues 2-3, 1 June 2000, Pages 173-179.