Research in Information Technology

Peter Eades





Two more extreme topics

Find a good topic

Narrow Nancy

The effect of the use of critical path planning in managing software projects

How to manage software projects

Broad Betty

Narrow and deep: An investigation of a few variable parameters, with many parameters held fixed.

Wide and shallow: Considers many parameters at once.

Narrow and broad

Find a good topic



Narrow Nancy

<u>Assume</u>

- an OO design method
- Java
- small teams
- 10K 100K SLOC
- Investigate effect of
- use of critical path planning

Broad Betty

Investigate the effects of

- 15 different design methodologies
- 7 different programming languages
- Small huge SLOC
- 17 different planning methods



Narrow and deep topic

<u>Advantages</u>

- More chance of pushing the boundary of knowledge
- More exciting

Disadvantages

- Your "model" may be too abstract and unrealistic
- It's hard to choose the variable parameters

Wide and shallow topic

Advantages

- Realistic
- Good training for industrial research

Disadvantages

- Mostly boring, like a collection of honours theses
- Unlikely to contribute a lot

Find a good topic Narrd My advice Choose a narrow and deep topic, and choose your variable and fixed parameters very carefully.

Another two extreme topics

Robustness theorems for nonpre-emptive scheduling methods

Fred the fundamentalist

Disk cache scheduling for Gnu C++ memory management on a Pentium 4 processor running Solaris

Andy the applicationist

Fundamental topic: abstraction of specific hardware and software Applied topic: specific hardware, specific software

Fundamental topic

Advantages

- Your thesis will have a longer life
- Your work can have more applications

Disadvantages

- It's hard to push the boundaries very far
- Your "model" may be too abstract and unrealistic

Applied topic

Advantages

- Easier problems
- May help with getting a job in industry
- Can contribute a lot to a relevant area

Disadvantages

- Your thesis can die young
- Your employment prospects can be shortlived
- Restricted applications

Popstar Paul

Another two extreme topics

P=NP

Classical Kirsty

Web-enabled distributed data mining for facebook social networks with 3D graphics

I want to solve an problem that has defeated many others

I want a lot of newspaper coverage

Classical topic

Advantages

- You may win the lottery and solve a hard problem
- Your thesis may have a long life
- Better referees
- Higher scientific quality

Disadvantages

- Can be frustrating
- Immediate rewards can be small

Hot topic

Advantages

- Better immediate feedback
- With good timing, you can get rich
- Easier to publish
- Easier problems
- Vibrant community
- **Disadvantages**
- Your thesis can die young
- Scientific quality can be low

er two extreme

<u>My advice</u>

Investigate a *fundamental* and *classical* topic, with some applications to a couple of *hot* and *applied* topics.

General advice on topics

Investigate a classical, fundamental, deep, and narrow topic, with some (perhaps shallow) applications to a couple of hot applied topics.
Obtain breadth by being a member of a team or research community

Also

Two extreme topics

Irene the introvert

2²³¹-1 is a prime number

This problem has been bothering me for decades. I can't rest until I know the answer.

Find a good topic

Eddie the extravert

2²³¹-1 is a prime number

A guy in a software security company has been phoning my supervisor to ask about this "possibly prime" number, **2**²³¹-**1**. I'll try to solve the problem.

Two extreme topics

Irene the introvert: selfmotivated, wants to find out for her own sake.

Eddie the extravert: Has a customer who wants to know, he will try to find out

There is no customer

Customer oriented

The customer may be an industrial partner, or a separate community of academic researchers

Introverted research

<u>Advantages</u>

 More exciting for some people

Disadvantages

- Funding unlikely
- May be worthless to everyone except yourself

Customer-oriented research

Advantages

- Good chance of good feedback
- Good chance of funding
- Better scientific criticism
- Better grounded in reality

Disadvantages

riented

<u>My advice</u> Ensure that you have *a customer*

Int

- The customer is possibly but not necessarily an industrial customer
- The customer may be another research group
- The customer should be outside your own research community
- The customer should be interested in results, not in methods

Thesis structure:

- Fundamental principles
- Case studies, some in the context of your customers
- Refer to case studies of other team members

```
PhD Thesis
```

```
- - -
```

Chapter 2 Investigation of a very difficult well known classical fundamental problem

. . .

. . .

Chapter 5: Case study 1, some hot topic Chapter 6: Case study 2, applied topic in customer context

```
. . .
```

Chapter 9 *Conclusions*: Refer to case studies by your colleagues

A model of optimizing compilers

- Chapter 2 Describe a new model for optimal code Chapter 3 Algorithms for creating optimal code under this model
- Chapter 5: Case study 1: how this model applies to mobile agents Chapter 6: Case study 2: comparison of optimal/sub-optimal code in a distributed transaction system

Chapter 9 Conclusions: further support for your hypotheses from work of your colleagues

Classical fundamental problem

Applications to hot and very applied topics

2. Do Research

Do Research

The research procedure

- 1. The customer has a problem.
- 2. The researcher produces an initial model of the problem.
- 3. <u>Repeat</u>
 - a) The researcher solves the problem, according to the model.
 - b) The researcher evaluates the solution of the model problem.
 - c) The customer evaluates the solution to the real problem.
 - d) The researcher adjusts the the model.

Until the customer is satisfied.

Do Research

Researchers have several roles to play

- 1. Create and adjust models of problems
 - ✓ abstract away non-essential details
 - use scientific theories and formalisms
- 2. <u>Solve</u> model problems
 - Use skills in CS/Math/sociology/psych/commonsense ..
 - ✓ Form hypotheses and solutions
- 3. <u>Evaluate</u> hypotheses and solutions to the model problems
 - ✓ Use skills in Math/Experiments/UCST

Create a model





1. Creating/adjusting a model

- A model is formed by *forgetting* some of the parameters of the real problem; models are simplifications of real problems.
- In IT, models are usually formal and mathematical.

Create a model

In practice, many models are models of models.



Example: the plotter problem

A pen plotter is a calligraphic device: it has a pen which moves over the paper to draw the picture.



The plotter problem

- A pen plotter has a pen which can be <u>up</u> or <u>down</u>.
- It accepts a sequence of penUp/Down/moveTo instructions.



penUp; moveTo (20,80) penDown; moveTo (80,80) penUp; moveTo (20,20) penDown; moveTo (80,20) penUp; moveTo (20,20) penDown; moveTo (80,20) penUp; moveTo (20,80) penUp; moveTo (80,80) penUp; moveTo (80,80) penUp; zero

The order of the instructions has an effect on the pen-up time.

The plotter problem:

Sort the instructions into an order that minimizes pen-up time.

penUp; moveTo (20,20) penDown; moveTo (20,80) moveTo (80,80) moveTo (80,20) moveTo (20,20) penUp; zero





Say speed = 5cm/sec

Bad solution: Pen-up time = 71 seconds.

Good solution: Pen-up time = 6 seconds

<u>The model</u>

We have:

- A set of "<u>primitives</u>"
- Each primitive has a <u>start point</u> and a <u>finish point</u>.
- The <u>pen-up time</u> is the sum of the distances from the finish point of one primitive to the start point of the next primitive.

We want:

An ordering for the primitives to minimize *pen-up* time.

The model *forgets* some parameters:

- The encoding system for the instructions
- The size of the paper
- The colours

Solving the problem

2. Finding a solution

Solutions are artifacts that help the customer.

Artifacts that make up a solution

Programs

Metaphors

Protocols

Architectures

Algorithms

Solving the problem



A solution is found using the skills of the researcher.

Your skill set is probably not enough to create a solution.

You probably need to increment your skill set

- Remember your undergraduate work
- Read books and research papers
- Attend seminars and conferences
- Ask your supervisor

Better research comes from a better skill set.

Solving the problem

Researchers draw on a number of fundamental skills to create a solution consisting of a number of artifacts.



Back to the plotter



<u>Solution</u>

One easy solution is the greedy algorithm:

- 1. Choose the first primitive so that its start point is the closest start point to PEN_ZERO.
- 2. <u>**Repeat**</u> for k=1 to NUM_PRIMS-1
 - Choose kth so that its start point is the closest unused start point to the previous finish point.

For example:

Draw an upper case "E"

Primitives:

- Line from (0.3, 0.2) to (0.3, 0.8)
- Line from (0.3, 0.2) to (0.7, 0.2)
- Line from (0.3, 0.5) to (0.7, 0.5)
- Line from (0.3, 0.8) to (0.7, 0.8)

The problem:

• Order these greedily


Greedy path:

- 1. Start at (0.0,0.0)
- 2. PenUp; move to (0.3, 0.2)
- 3. PenDown; move to (0.3, 0.8)
- 4. PenDown; move to (0.7, 0.8)
- 5. PenUp; move to (0.3,0.5)
- 6. PenDown; move to (0.7, 0.5)
- 7. PenUp; move to (0.3, 0.2)
- 8. PenDown; move to (0.7, 0.2)
- 9. PenUp; move to (0.0,0.0)



If PenUp moves are (*xstart_i*,*ystart_i*) to (*xfinish_i*,*yfinish_i*) then we can calculate the total PenUp time by adding up the Euclidean distances:

$$PenUp = \sum_{i} \sqrt{(xstart_{i} - xfinish_{i})^{2} + (ystart_{i} - yfinish_{i})^{2}}$$

In this case
$$PenUp = \sqrt{0.13} + \sqrt{0.25} + \sqrt{0.25} + \sqrt{0.51}$$
$$= 2.074$$



3. Evaluating a solution

To evaluate a solution, you need

- An evaluation <u>measure</u> that tells you whether the solution is good or bad
- An evaluation <u>method</u> to compute the measure

Evaluation measures

There are three basic <u>measures</u> for the quality of a solution:



The E³ measures:

- Effectiveness: is the solution logically correct? Is it optimal? Is it satisfactory for the customer?
- 2. <u>Efficiency</u>: does the solution use computational resources efficiently?
- **3.** <u>Elegance</u>: is the solution beautiful, simple, and elegant?

My conjecture: All solutions can be measured in terms of these three parameters.

There are three basic *evaluation methods*



And many combinations of these approaches

The three methods:

- 1. Mathematics
 - You prove a theorem that says that the solution is effective/elegant/efficient
- 2. Experiments
 - Run programs on test data
 - Test systems with human subjects
- 3. UCST: Try to sell your solution

My conjecture: These are the only evaluation methods in information technology.

<u>Evaluation of the greedy plotter optimisation by</u> <u>UCST</u>

1. The greedy solution can be "proven" *effective* by UCST.

UCS Assertion:

"Since it chooses the best alternative at each stage, it gives minimum pen up time".

This may be convincing for some customers, but not for PhD thesis examiners.

2. The greedy solution is *elegant* by UCST: it is easy to understand, easy to implement.

Mathematical Evaluation

The greedy solution can be investigated for effectiveness using Mathematics.

a) Negative result: Greedy does not always give optimal results.



The optimal path is shorter.



Mathematical evaluation

b) The greedy method is *close* to optimal:

<u>Theorem</u>

If **GREED** is the pen-up time with the greedy solution and **OPT** is the pen-up time with the optimum solution then **GREED / OPT = O(logn).**

Proof: lots of mathematics and lots of complex mathematics lots of even worse mathematics and lots and lots of mathematics and even more complicated mathematics and more and an incredible amount of mathematics and lots of mathematics lots of mathematics and lots of mathematics lots of complicated mathematics lots of mathematics lots of mathematics and of mathematics lots of complicated mathematics lots of mathematics lots of mathematics and of mathematics lots of complicated mathematics lots of mathematics lots of mathematics and of mathematics lots of complicated mathematics lots of mathematics lots of mathematics and of mathematics lots of complicated mathematics lots of mathematics lots of mathematics and of mathematics lots of complicated mathematics lots of mathematics lots of mathematics and

Experimental Evaluation





Experiments showed that greedy is very close to optimal: for larger plots it is within 10% of optimal.

BUT . . .

Experimental Evaluation

We replaced the quality evaluation with a real plotter



And timed the real plotter using the wall clock. The customer was happy, but it revealed two problems:

- a) The model was wrong,
- b) The greedy algorithm was too expensive.

The research procedure

- 1. The customer has a problem.
- 2. The researcher produces an initial model of the problem.
- 3. <u>Repeat</u>
 - a) The researcher solves the problem, according to the model.
 - b) The researcher evaluates the solution of the model problem.
 - c) The customer evaluates the solution to the real problem.
 - d) The researcher adjusts the the model.

Until the customer is satisfied.

a) <u>Our model was wrong</u>

At a micro-level, the plotter pen moved in three ways:

- Horizontally
- Vertically
- (some plotters) At 45° to horizontal

Each micro-movement takes one unit of time. This implies that the distance function is L^{∞} rather

than L^2 .





Mathematical Evaluation

It was easy to check that the mathematical results remain true for any distance function, and this change in model did not change the theorems significantly.

Experimental Evaluation

We repeated the experiments and the results changed a little, but the general pattern was the same.

b) Our solution was not efficient



The greedy algorithm runs in time $O(n^2)$. This was slower than the drawing procedure.

Solution: optimize one buffer-sized section at a time.



An "optimized" bufferful is sent from the greedy algorithm to the buffer whenever the plotter exhausted the current buffer.

The bufferised greedy algorithm was almost as effective as the straight greedy algorithm, and much faster.



Lessons from the plotter problem

Mathematics	 Robust to model changes Good evaluation of pathological behavior 	 Does not evaluate the model
Experiments	 Evaluates the model Good evaluation of normal behavior 	 Poor evaluator for pathological behavior
UCST	 Convinces the non- scientific customer OK to evaluate elegance 	•Poor evaluator of efficiency / effectiveness.

Do Research

<u>My advice</u>

To find a solution:

- Use your own skills
- Read a lot
- Attend seminars and conferences *To evaluate your solution*
- Concentrate on mathematical and experimental methods, avoid UCST
- Relate your results to E³: effectiveness, efficiency and elegance

- 3. Present results
 - a. Write papers
 - b. Give talks
 - c. Write a thesis

3a: Write papers

You can write

- Papers in NLCs
 - ✓ nice local conferences
- Papers in IK-CCs
 - ✓ international killer-competitive conferences
 - ✓ Rated A or A⁺
- Papers in journals
- Chapters in books
- Books

Students mostly write conference papers; I will concentrate on this.



How the process works

- a) You write the paper
- b) You submit the paper to the program committee chair
- c) The program committee chair sends it to members of the program committee (takes about a week)
- d) They read it (in about 4 weeks) and write a brief report. They decide whether to accept your paper
- e) If your paper is accepted, you revise the paper according to the referee's comments (2 4 weeks)
- f) You give a talk at the conference

<u>How do the program committee decide</u> which papers to accept?

- In most cases, the papers are scored and sorted on score.
- Very few papers get a very high score or very low score.
- Accept/reject decisions for middlescore papers can be fairly arbitrary

10 - 20% Obviously Accepted	60 - 80% random and ad-hoc decisions	10 - 20% Obviously Rejected
-----------------------------------	---	-----------------------------------

Three steps



2. Choose a good conference, and adjust your paper to that conference

3. Send the paper, sit around and hope that it is accepted



<u>2. Choose a good conference, and adjust your</u> paper to that conference

Choose a conference

- ✓ The best conference possible (A or A⁺)
- ✓ A good program committee
- ✓ Realistic deadline
- ✓ Avoid "scams"

Adjust your paper

- Motivation aimed toward the conference community
- Research methods that are familiar to the conference community
- Don't insult people on the program committee

<u>3. Send the paper, sit around and hope that</u> <u>it is accepted</u>

• Don't worry if it is rejected.

How to get your paper rejected

The top methods

- 1. Write in bad English
- 2. Be unaware of current trends in the specific conference community
- 3. Organize your thoughts badly
- 4. Omit motivation

Extending the three steps to write a journal paper:



Give good talks

3b. Give lots of good talks

Give good talks

<u>Giving a talk is beneficial to the</u> <u>speaker</u>

- It helps you
 - ✓ define your problem
 - ✓ understand your own work
 - ✓ organize your ideas
 - ✓ become famous
 - \checkmark write a thesis
- It brings feedback from others


You should have three talks ready to give at any time:

- a) 30 minute talk/demo
 - ✓ For a conference, …
- b) 5 minute talk/demo
 - ✓ For a research visitor, at a poster session, …
- c) 30 second explanation of what your research is about
 - ✓ For when you are in the elevator, …

How to give a talk at a conference

Giving a talk consists of three elements:

- a) Organization
- b) Talking and walking
- c) Visuals

These elements vary depending on the type of presentation.

Some comments about research conference presentations . . .

a) **Organization**



Example:

Title: Fast spatial data mining in low dimensions

OData mining helps people 5 Your data mining algorithms: Everyone description at a high level understands • no proofs, no details 15 Some Proof of the 2D case understand 20 Chart of experimental results 23 Repeat main results 25

b) <u>Talking and walking</u>

- Look at the audience as much as possible
 Choose specific people to focus on
- Speak slowly and clearly, and avoid idiomatic English
 ✓ English is a second language to most people in IT
- Use your hands for expression
 ✓ avoid holding a microphone
- Don't waste time
 - Check your data-projector/laptop connection
 - ✓ Have your ppt well sorted out before you start

c) <u>Visuals</u>

- Use a medium that is suitable
 - ✓Use a computer for graphics
 - ✓Use a blackboard for mathematics
- Use a medium that is well supported by the local system
- Ensure that your visuals are perfect
 - ✓No speeling errors
 - ✓No spacing errors
 - Attractive layout (e.g., avoid linebreaks as much as possible)
- Don't use visuals as notes to yours
- Use pictures wherever possible
- Avoid ducks



Look at the audience; avoid ducks



Look at the audience; use your hands



Look at the audience; avoid holding a microphone; ensure that your slides are perfect





Ensure that your slides are perfect

In this paper, based on statistical and visual similarity analysis for the correlation among multi-spectral planes, a novel approach to multi-spectral image compression is proposed. The algorithm differs from classical lossy approaches of multi-spectral image coding in the fact that it provides the further decorrelation scheme of the spectral planes by inter-plane transformations and coding, which is based on information distribution. Moreover, the human visual system is introduced into the transformation. In the process of doing so, it will play a very active role in fully exploring the psycho visual redundancy. The new technique for multi-spectral image compression, which is designed to be compatible with the JPEG standard, is demonstrated on extracting correlation among planes based on human visual system. A high measure of compactness in the data representation and compression can be seen with the power of the scheme taken into

Use the slides for the audience, not as reminders for you

Formal specification of Security Protocols

- The need for security
- The need for formal specification
- Porter and Quirk's language
- Inadequacies

More advice

- Give a practice talk to your team
- Ask people to look out for
 - errors and ducks in the visuals
 - idiomatic and ambiguous English
 - not looking at the audience and write it all down, and tell you
- Video the talk, look at the video

It is very important to write a good thesis.

Your 3⁺ years of PhD research are examined on the basis of:

- 1. your thesis.
- 2. your thesis.
- 3. your thesis.

Not on the basis of

- Computer systems that you have coded
- Undergraduate tutorials that you have given
- Ideas that you have had

The examiner reads your thesis, and not much else, then writes a very simple report.

				13. [±] ār									2
										4	The th submi	esis does t the thes	not meet the standard expected for the degree but the candidate be permitted to r is for examination to the examiners after -
PHD EXAMINERS' SUMMARY REPORT FORM												4.1	re-writing one or more sections of the thesis in light of the examiners commer specified in the General Report Form
				(For candidates enrolled in the degree of PhD by thesis)								4.2	undertaking further work and revising the thesis to reflect the additional work
Candidate's Name:								OB			4.3	in addition to 4.1 or 4.2 above, presenting for an oral examination	
Thesis Title:									5	The c	indidate s	should be considered for the award of a masters degree	
Examiner: Professor Peter Eades				(E	xaminer No. 1)						5.1	without the requirement for revision or further examination	
							-				ũ	5.2	subject to revisions specified in the General Report Sheet being completed to the
For info Guidanc	ion abo Examin	out the sta ners of Do	andard for the award of the degree please consult the attached octoral Submissions'.	documen	t, 'Notes for the						61	saustaction of the Champerson of Examiners	
Plase answer the following questions about the thesis. You are asked to provide more detailed comments on the one of the following questions about the thesis. You are asked to provide more detailed comments on the one of the following questions about the thesis.									Report Form				
accompanying General Report Form.									ld be rejected, and the degree not be awarded				
THE T	IESI	IS											
2 3 4 5	Does the thesis make an original and significant contribution to knowledge If yes and understanding of the field of study with which it is concerned? Is the standard of litrary presentation in the thesis satisfactory? If yes appropriate for the thesis topic and the degree sought? Does the thesis subble for publication as a book or in a learned journal S.1 in the form submitted? S.2 with modifications? Different set in the summation is finalised.												
RECON	RECOMMENDATION ON RESULTS OF EXAMINATION								2	8/	3 .2	01	SIGNATURE:
C	ב	I The candidate should be awarded the degree without the requirement for revision or further examination POSITION: Latesser DISTITUTION: University, of Sydney						INSTITUTION: University of Sydney.					
0	7	2	Subject satisfacti	ubject to minor revisions specified in the General Report Form being completed to the attifaction of the Chairperson of Examiners, the candidate should be awarded the degree									
C		3	The candidate be awarded the degree subject to the requirement that he/abe present for an oral defence										
PLEASE TURN THE PAGE													
C	R												
Research and Postgraduate Studies Committee (Jan 2000)									h and P	ostgradu	ate Studie	Committe	e (Jan 2000)

Plus three or four pages of comments . . .

Your examiner basically just ticks a box

	1	The candidate should be awarded the degree without the requirement for revision or further examination										
	2	Subject to minor revisions specified in the General Report Form being completed to the satisfaction of the Chairperson of Examiners, the candidate should be awarded the degree										
	3	The candidate be awarded the degree subject to the requirement that he/she present for an oral defence										
	4	The thesis does not meet the standard expected for the degree but the candidate be permitted to re- submit the thesis for examination to the examiners after -										
			4.1	re-writing one or more sections of the thesis in light of the examiners comments specified in the General Report Form								
		OR OR OR	4.2	undertaking further work and revising the thesis to reflect the additional work								
OR			4.3	in addition to 4.1 or 4.2 above, presenting for an oral examination								
	5	The candidate should be considered for the award of a masters degree -										
		D OR	5.1	without the requirement for revision or further examination								
			5.2	subject to revisions specified in the General Report Sheet being completed to the satisfaction of the Chairperson of Examiners								
		OR	5.3	subject to re-examination after completion of revisions specified in the General Report Form								
	6	The thesis should be rejected, and the degree not be awarded										

Your examiner basically just ticks a box

1. Award a PhD

- 2. Award it after some minor corrections
- 3. Award it as long as the student makes some corrections
- 4. Ask the student to rewrite part (or all), and re-submit
- 5. Tell the student to go away.

Evaluation of a thesis

Examiners are basically asked: *"Is this a good thesis?"*

The evaluation measures vary from one University to another.

- Some typical measures:
 - Original and significant contributions
 - Methodology
 - Expression
 - Scholarship, reference to literature

<u>Length</u>

The research content of a thesis should be about 3 good journal papers.

However, a thesis is different from a paper

- It has to tell a single story
- More background
- More references
- Extensive evidence of all the claims
- Justification of the research methodology

<u>My advice</u>: before you begin to write:

- Carefully read at least one thesis from someone outside your field.
- Read at least 3 examiners reports

My advice: The writing process

- Take 3 4 months
- Write about 150 pages; about 3 pages per day for the first draft
- Ensure that your supervisor reads every word
- Get someone outside your field to read the introduction
- List your original contributions in the first chapter

How to get your thesis rejected

Some top methods

- 1. Take a job before you submit
- 2. Teach more than one undergraduate unit
- 3. Don't evaluate your solutions
- 4. Ignore feedback
- 5. Be unaware of current trends in your research community
- 6. Organize your thoughts badly
- 7. Write a very long thesis

Winding up now ...

<u>Topics that I have not mentioned</u>

- 1. Part-time or full-time?
- 2. Managing your time
- 3. Three stages of a PhD candidature
 - a) Learning
 - b) Research
 - c) Writing
- 4. What do you do when something goes wrong?
- 5. PhDs and careers
 - a) What kind of PhD leads to an industrial career?
 - b) What kind of PhD leads to an academic career?

Conclusion

- 1. Find a good topic
- 2. Do Research
- 3. Present your research
 - a) Give lots of good talks
 - b) Write lots of good papers
 - c) Write a good thesis
 - 4. Have fun . . .