CSE Undergraduate Theses - Introduction

What’s going on ….? How to succeed … ?

Presented by Helen Paik

CSE Thesis Team

People who can help:

- **Thesis Supervisor** … all thesis related issues
- **Thesis Coordinator** ... Helen Paik
  - ‘other’ thesis related issues: nominations, guidance, staff liaison
  - Email: hpaik@cse Room: K17-401A, by appointment
  - Let’s utilise the forum function of the new thesis site (in WebCMS3)
- **Student Office ...**
  - administration: extensions, late penalties, reassessment
  - Email: office.thesis@cse, Room: K17-Office, Office opening hours

Thesis Intro Lecture

The goals of this lecture:

- describe the process and requirements for a CSE thesis
- get you started with Thesis A …

At the end, you should ...

- understand what's required of you
- start to plan your time as a thesis student

Aims of 4th-year Thesis

- allow you to "**put together**" what you've learned
- give you experience in tackling a **sizeable** project
- give you exposure to **research/implementation** topics
- require you to practice **planning/time-management**
- give you experience in **formal report writing and presentation**

Overview of Thesis (as-a-whole)

- find a topic (do this now)
- do background research, make plan (S2, weeks 1-7)
- seminar presentation (project overview + plan) (S2, week 7)
  - follow-up on feedback, start work on solution ...
- report (literature review + plan) (due S2, week 12)
  - keep working on solution, evaluate results ...
- final presentation/demonstration (next year S1, week 11)
- thesis (project + evaluation) (next year due S1, week 13)

Deliverables in Thesis A

Thesis Seminar Presentation (during Week 7) (30%):

- a 30 minute presentation about your Thesis A topic and the plan
- organise the time/date with School Office

Thesis A Written Report (Tuesday 11:59pm Week 12) (70%)

- Literature review + design and implementation plan of your thesis
- Use the template provided (thesis course web site)

Thesis Seminar Attendance Sheet (during Week 7):

- Attend 4 (four) seminar presentations of other Thesis A students.
- Submit the attendance sheet with signatures
- No mark, but a requirement for passing Thesis A.

Deliverables in Thesis B

Thesis Demonstration/Presentation (During Week 11) (20%):

- a 30 minute presentation about the final outcome of your thesis.
- Organise the time/date yourself with supervisor+assessor

Final Thesis Report (Tuesday 11:59pm Week 13) (80%)

- This is “the thesis” (i.e., should read like a whole piece by itself)

Thesis Summary/Abstract (Week 13):

- you are required to submit 150-word summary of your thesis (besides the report).

Draft Manuscript (Week 14) - for COMP4941

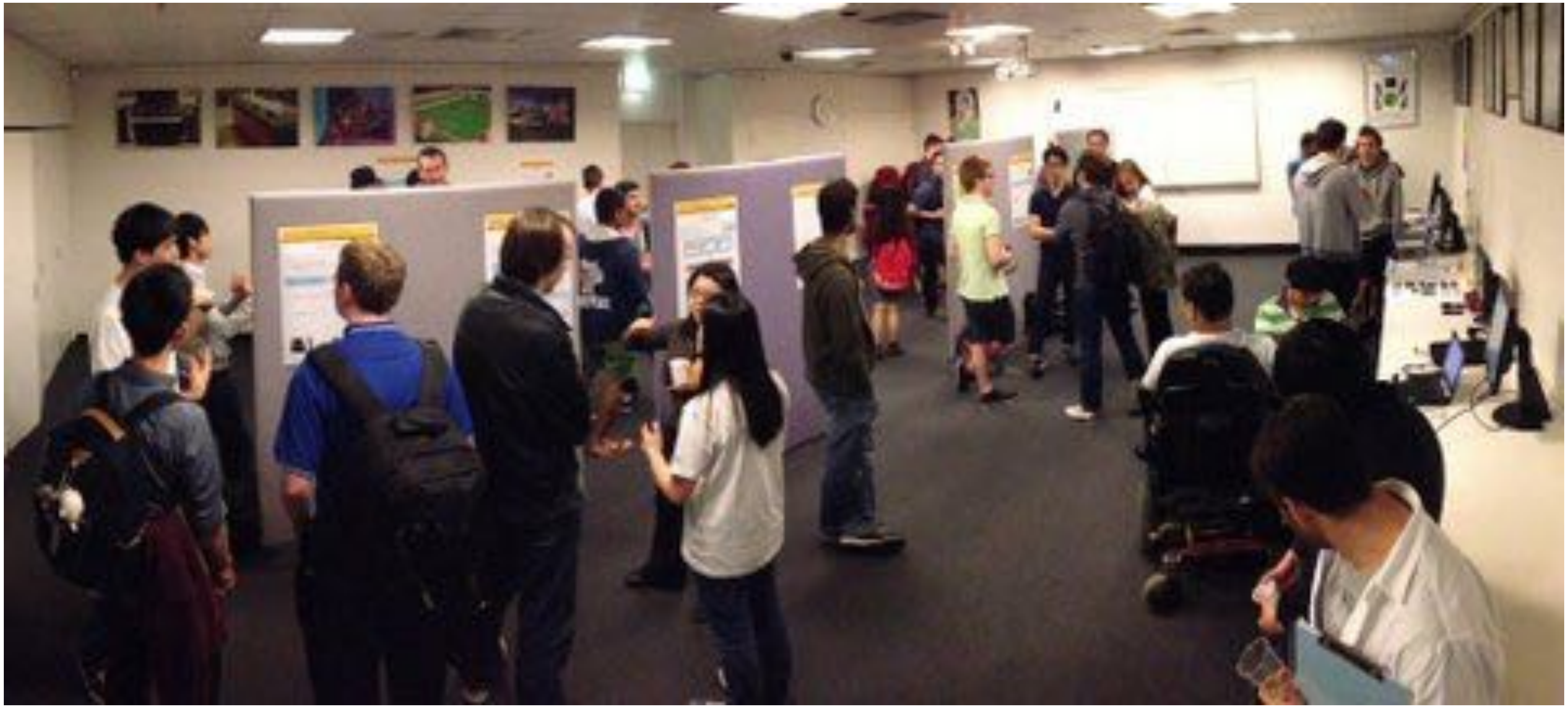
Thesis Showcase

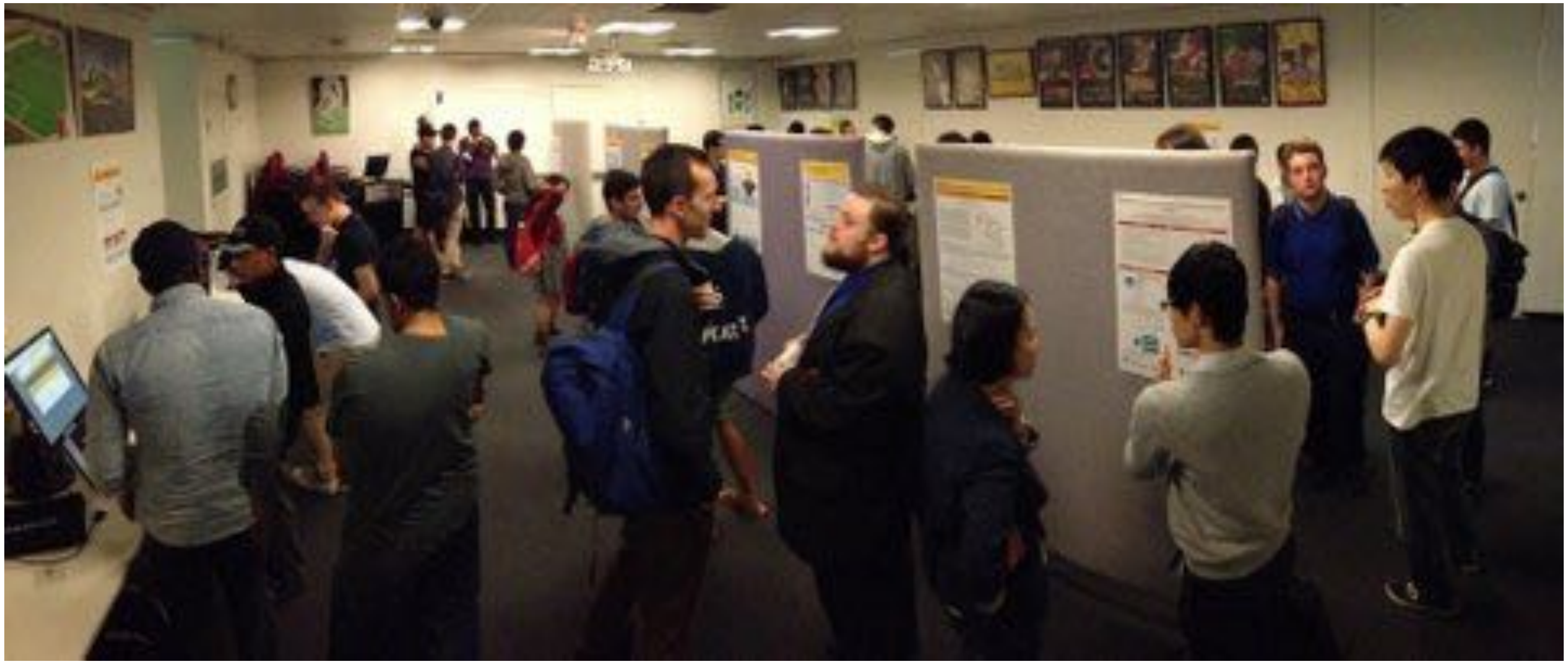
Next year S1 Week 13 Thursday 5.30pm

Students with good demo results will be invited

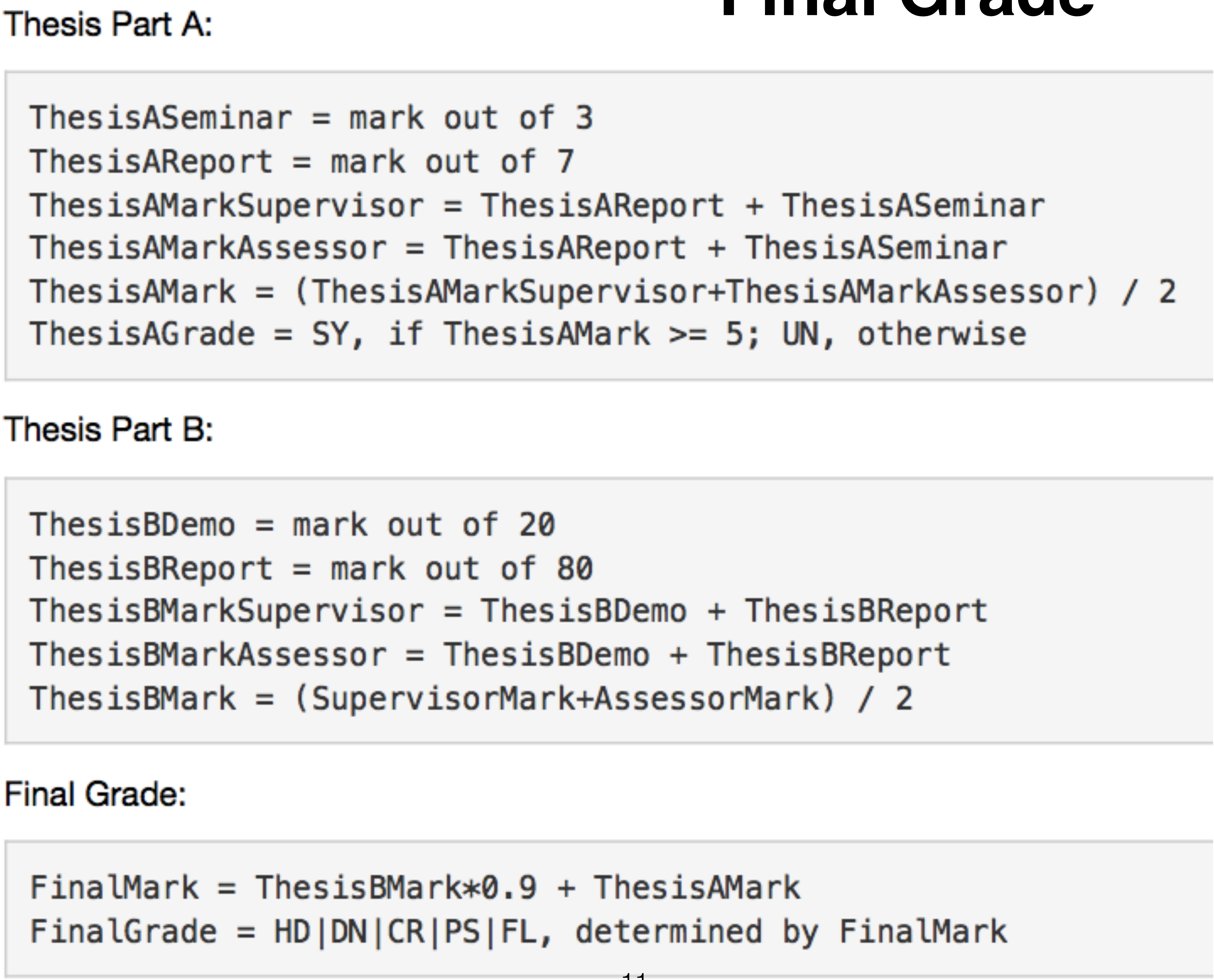
- poster/demonstration sessions.
- guests: year 2/3 students, staff members, industry sponsors, postgrad students
- **food (not pizza)**

Need to submit a poster (extra work), but worthy event to participate





Final Grade



Thesis Part A:

```
ThesisASeminar = mark out of 3
ThesisAReport = mark out of 7
ThesisAMarkSupervisor = ThesisAReport + ThesisASeminar
ThesisAMarkAssessor = ThesisAReport + ThesisASeminar
ThesisAMark = (ThesisAMarkSupervisor+ThesisAMarkAssessor) / 2
ThesisAGrade = SY, if ThesisAMark >= 5; UN, otherwise
```

Thesis Part B:

```
ThesisBDemo = mark out of 20
ThesisBReport = mark out of 80
ThesisBMarkSupervisor = ThesisBDemo + ThesisBReport
ThesisBMarkAssessor = ThesisBDemo + ThesisBReport
ThesisBMark = (SupervisorMark+AssessorMark) / 2
```

Final Grade:

```
FinalMark = ThesisBMark*0.9 + ThesisAMark
FinalGrade = HD|DN|CR|PS|FL, determined by FinalMark
```

FAQ

- Q: How long should X be? (X ∈Chapter,Report,Seminar,Thesis)
  - A: As long as is necessary to make it convincing.
- Q: When is Y due? (Y ∈ Report,Seminar,Thesis)
  - A: Check the thesis course home page.
  - A: Help fellow Thesis A students out through the course forum for any ‘how to’/‘where to’ questions …
- Q: How much time should I spend on my Thesis?
  - A: Notionally, 150 hours per 6UoC - roughly 10-12 hours a week. But, generally, the more time you spend, the better the outcome

FAQ

- Q: What happens if I can't finish?
  - A: You get less marks than you would if you finished. (The definition of "finished" is looser for thesis than assignment)
- Q: Can I get an extension?
- A: as per usual special consideration procedure
- A: If Thesis A is missing, you get AF. If late, zero (0)
- A: If Thesis B is late, you suffer heavy late penalties (5 marks per day off the thesis report mark).

FAQ

- Q: What must I do to get good marks?
  - A: Depends on who you're asking ...
    - Supervisor: knows everything you did. May assess based on continuous performance
    - Assessor: (most likely) sees only Seminar, Demo, Thesis. Likely to assess based on what she/he observes in S/D/T
  - To be safe: ask what they're looking for in a good thesis

Why a Thesis is not an Assignment

- A thesis is significantly different from an assignment:
  - it is, typically, open-ended
    - there is not an obvious "correct" answer or end-point
    - you have more say in the direction the work takes
  - it has a much longer time-frame
    - you need more self-discipline to get things done
    - you have more responsibility to plan your progress
- Think about employing popular time management skills/tools

Different Types of Theses

- Theses have been classified into:
  - **RES** carry out a small focused piece of research
  - **DEV** build a software and/or hardware system
  - **R&D** combination of the above two ... build a system, but needs research to get it done
- Expectations for each type are slightly different

Research in Computing

- Writing a piece of software, no matter how complex, isn't generally regarded as research in itself.

- However, it would be considered research if

    - it uses a new method/algorithm/data structure

    - it solves a problem not previously solved by computer (e.g., new framework/platform)

    - it applies a solution from one area to solve problems in another area

Abstract

Integrating task and motion planning is a computational difficult problem which has a many of interesting applications in the real world. There are several ways to integrate task and motion planning, but usually authors take the approach to use motion planners as the core of their development. This approach have been deeply studied over the years and achieved useful results [Cambon et al. 2009; Plaku and Hager 2010]. However, the problem is far from being solved. The different characteristics of the task planner (discrete variables, finite search space) and motion planners (continuous variables, infinite configuration space) makes them hard to combine and even more complicated to get a sound and complete solution with great performance. The recent work by [Ivankovic et al. 2014] extends classical planning to integrate constraints given to an external solver. Here we adapt this work to integrate task planning with the motion planner OpenRAVE, which plays the role of the external solver. This Thesis attempts to correctly formalize, implement, and simulate the combined task and motion planning problem.

Research in Computing

- **the new solution must be demonstrably better than earlier approaches**
- Evaluation of computing research:
    - solves existing problem more effectively than before
    - solves a wider range of problems than before   (generalises)
- Demonstrations of effectiveness follow two tracks ...
    - Theoretical, e.g. analyse complexity, prove upper/lower bounds, ...
    - Experimental, e.g. build prototype; measure performance on range of data …
- **Make sure that you have a good conversation with your supervisor about this …**

Development in Computing

- Aim is to build a system to meet a demand or solve a problem.
- May involve developing software, hardware, or a combination.
- The goal is clearly to build the system, but you must also:
    - follow a (software) engineering methodology (+ document it)
    - provide a demonstration that the system works effectively
    - note any unsolved problems and limitations

One of the most serious dangers of a fall is the potential for a long period spent on the ground after a fall, often caused by a delayed discovery of the fallen person. Many older fallers are unable to get up again without assistance, and that subsequent long lie can lead to issues such as muscle damage, hypothermia, and dehydration.

Dedicated devices that can detect falls to a high degree of accuracy are available, but these devices are usually only worn by those at high risk of falling. With the prevalence of smartphones in todays society, and the motion sensors that are integrated into them, a smartphone can potentially achieve the same fall detection accuracy using a device that the majority of people carry with them anyhow.

As technology improves, manufacturers are able to include more sensors in smartphones. Some phones now include a barometric pressure sensor, which can sense small changes in atmospheric pressure, and give an indication of vertical movement of the phone. Using this in combination with existing accelerometry-based techniques could lead to an even more reliable fall detection device that is also more convenient to use.

Ethics Clearance

- https://research.unsw.edu.au/human-research-ethics-home

Doing Thesis A

- Thesis A aims for you to demonstrate that ...
    - you have a thorough understanding of the topic
    - you have identified an area that requires work
- you have an approach for solving the problem
- you have a plan to demonstrate the likely effectiveness of this approach
- you have a plan for carrying out the work (including time-frames for tasks)

Doing Thesis A

- Specific tasks for Thesis A ...
  - accumulate a collection of references that
    - discuss issues related to the problem being addressed
    - describe attempts by others at solving the problem
  - describe/analyse the problem (aided by references)
  - for DEV theses: produce detailed requirements/spec
  - establish an evaluation framework; analyse prior work
    - consider Ethics Clearance
  - draw up a plan for work to solve the problem
  - start work on solving the problem

Doing Thesis A

- Suggested timetable for Thesis A work:

Weeks	Task
1-3	Meet your supervisor, and sort out what your project is
2-10	Collect and read relevant literature
2-10	Make notes on your reading (build them up gradually)
4-7	Prepare seminar presentation (Seminar)
4-8	Prepare/revise your method and plan
4-11	Write the report
12	Report
ASAP	Start working on solution

Thesis A Seminar

- Typical Seminar Structure
  - Introduction: sell the topic, summarise aims
  - Background: set context, evaluate previous work
  - Proposal/Plan: how do you plan to tackle the problem
  - Bibliography: give references for all work cited
- Seminar = summary of Report, publicity for project, chance to get feedback

Thesis A Seminar

- The seminar aims to:
  - give you a chance to practice your presentation skills
  - let you show that you have met the goals of Thesis A
    - convince others that you're studying an important/interesting problem
    - demonstrate that you've done some research/thinking about it already
    - have a plan for the rest of the year to solve the problem
- If you already have some results to show, that's a bonus.
- Target your seminar at fellow thesis students (general audience)
- Target the hard-core technical stuff at your supervisor and assessor.

Thesis A Seminar

- 45 minute timeslot is allocated for each presentation:
  - 25 mins talk, 5-10 mins Q+A with audience
  - 10-15 mins debrief with supervisor/assessor
- Take it seriously ... you're being assessed.
- Use max 20 slides; you cannot cover more in 25 mins.
- Pay attention to questions - good source of feedback/ideas
- Attend other people's seminars (requirement).
  - you might get some ideas for your own project
  - they get a chance to present to an audience

Thesis A Report

- Typical Thesis A Structure:
  - Introduction: sell the topic, summarise aims (1-2 pages)
  - Background: set context, evaluate previous work (4-6 pages)
  - Proposal/Plan: how do you plan to tackle the problem (with justification based on ideas in Background) (6-8 pages)
  - Bibliography: give references for all work cited (1-2 pages)
- "set context" = define/examine problem in detail, set out evaluation framework

Academic Writing Style

- Thesis/report both have overall structure:
  - Introduction ... what the thesis is about
  - Main Part ... the details of the work
  - Conclusion ... what the thesis achieved
- Individual chapters should follow a similar structure:
  - Introduction ... what this chapter is about
  - Main Part ... the details of the chapter
  - Summary ... what the chapter achieved
- May sound repetitive but it provides linkage and rationale for the reader.
- Use **the thesis template** provided (LaTeX and Word)

Academic Writing Style

- UNSW Student Resources (+ many other university online resources)
- A few common (easy) tips that you can immediately use:
    - Try to be “formal”, “technical”, “impersonal”
        - Using “I” ?
        - Don’t, Isn’t ?
        - “a bit”,“get used in …”?
        - “wonderful”, “beautiful”, “terrible”, “hopeless”, “useless”, “amazing”, etc.
    - Introduce and define “terms” properly before start using them
    - Introduce acronym properly when first used
        - e.g., The University of New South Wales (UNSW) is …
    - Use caution: This may cause, vs. I think this must cause, or say ``There is evidence to support that …’’
    - Use Active Voice whenever possible

Doing the Literature Review

- Goals:
  - collect a comprehensive set of publications on the topic
  - build a picture of the nature and scope of the problem
  - develop a framework for evaluating possible solutions
  - analyse the specific work described in the publications
- How comprehensive? (a.k.a. how many references?)
  - until you are convinced that you have all relevant materials
  - use your judgement when to stop (and **ask supervisor**)

Doing the Literature Review

- Some tips ...
  - try to identify seminal papers on the topic (ask Supervisor)
  - use bibliographies to find prior work
  - use Citation Index to find subsequent work (e.g. Google Scholar)
  - maintain a database using a bibliography tool (e.g. Mendeley)
  - read and think about the references
  - keep electronic notes; describe in your own words
  - identify common themes, structures and assumptions

Orchestration

have been found and bound
hestrate them to execute their
. Fig. 3 shows the lifecycle of
ution consists of:

stage is to call the bound
luce data for the task's out-
:h provider is automatically
ponse comes asynchronously
create or update the output.
e multiple responses to one
pdate, and the request can be
iders'response.

to evaluate the task's output
by the bound provider. Ac-
agreement, evaluation can be
ıe requester), or automatically
ıe third-party service). Based
and detailed agreement terms,
l or redone.

ed as a set of interdependent
may drive another, based on
e consider the interdependen-
that can be expressed by an

References

[1] D. C. Brabham, "Crowdsourcing as a model for problem solving an introduction and cases," *Convergence: the international journal of research into new media technologies*, vol. 14, no. 1, pp. 75–90, 2008.

[2] M. Vukovic, "Crowdsourcing for enterprises," in *SERVICES 2009 - 5th 2009 World Congress on Services, September 21, 2009 - September 25, 2009*, ser. SERVICES 2009 - 5th 2009 World Congress on Services. Bangalore, India: IEEE Computer Society, 2009, pp. 686–692.

[3] A. Kittur, J. V. Nickerson, M. S. Bernstein, E. M. Gerber, A. Shaw, J. Zimmerman, M. Lease, and J. J. Horton, "The future of crowd work," in *2013 2nd ACM Conference on Computer Supported Cooperative Work, CSCW 2013, February 23, 2013 - February 27, 2013*, ser. Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW. San Antonio, TX, United states: Association for Computing Machinery, 2013, pp. 1301–1317.

[4] M. P. Papazoglou, P. Traverso, S. Dustdar, and F. Leymann, "Service-oriented computing: state of the art and research challenges," *Computer*, vol. 40, no. 11, pp. 38–45, 2007.

[5] A. Kittur, B. Smus, S. Khamkar, and R. E. Kraut, "Crowdforge: Crowdsourcing complex work," in *24th Annual ACM Symposium on User Interface Software and Technology, UIST'11, October 16, 2011 - October 19, 2011*, ser. UIST'11 - Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. Santa Barbara, CA, United states: Association for Computing Machinery, 2011, pp. 43–52.

[6] A. Kulkarni, M. Can, and B. Hartmann, "Collaboratively crowdsourcing workflows with turkomatic," in *ACM 2012 Conference on Computer Supported Cooperative Work, CSCW'12, February 11, 2012 - February 15, 2012*, ser. Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW. Seattle, WA, United states: Association for Computing Machinery, 2012, pp. 1003–1012.

Using the References

- What you should NOT do with references:
  - copy/paste chunks of text from them into your report
  - if you do this, it's plagiarism and you fail
- Every statement in your thesis which is based on others' work
  - must be attributed to them (via a reference)
  - even if you make the statement entirely in your own words
  - but especially if you are "quoting" them (minimise this)

Using the References

M-trees do not assist Z queries. Even if the Z queries conform to the normal pattern of querying expected in this context, the algorithmic complexity is still too high

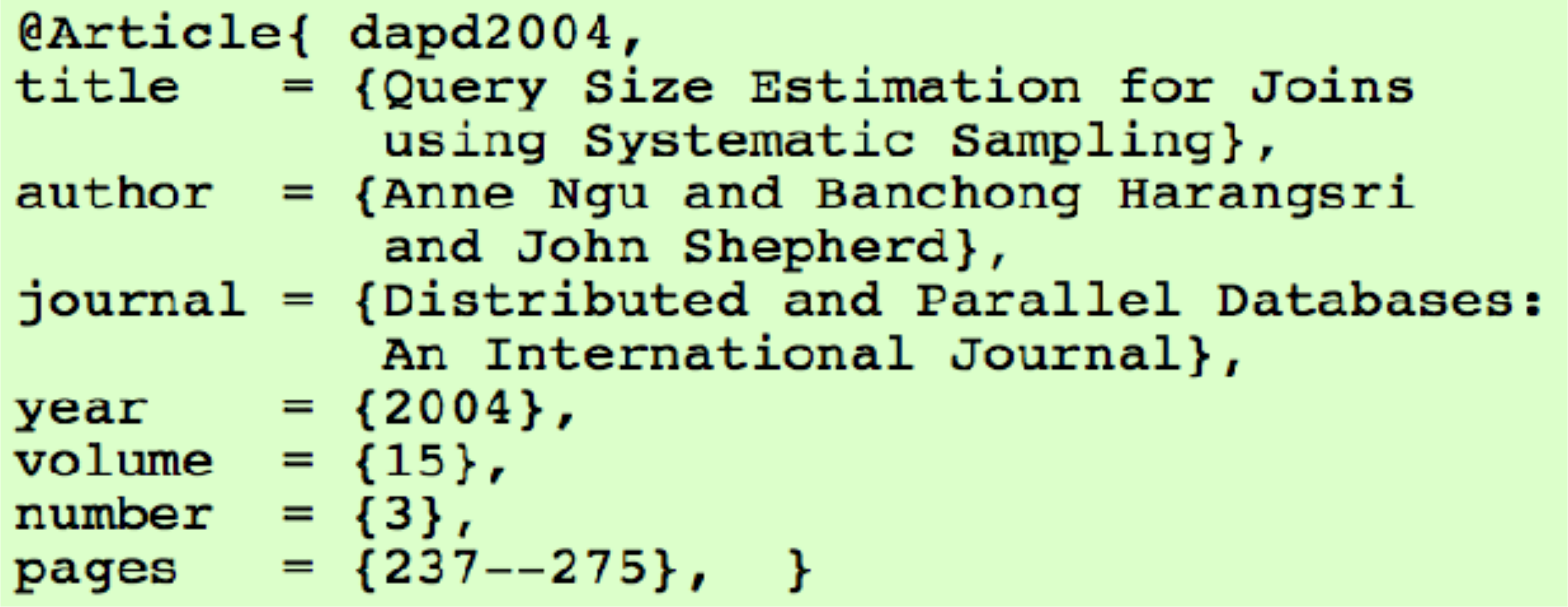
- Examples of acceptable use of others' material:
    - and Smith [8] noted "M-trees do not assist Z queries".
    - and (Smith, 1998) noted "M-trees do not assist Z queries".
    - ... and as was pointed out by Smith [8]:

The Bibliography

- The bibliography
  - consists of a list of all of references used in the report
  - with **enough detail** that a reader could find each reference
- It should **not be** simply a list of URLs.
- For each reference, there must be:
  - an author, a title, a date
  - information to identify publication source
- BibTeX has well-defined styles for different kinds of references.

The Bibliography

Example: BibTeX and reference for a journal article:



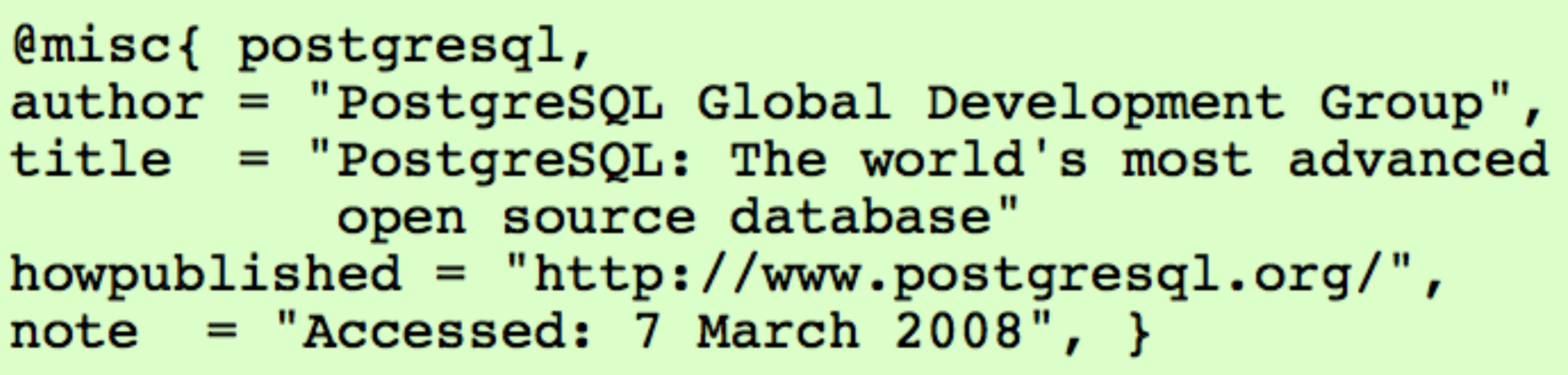
```
@Article{ dapd2004,
title    = {Query Size Estimation for Joins
           using Systematic Sampling},
author   = {Anne Ngu and Banchong Harangsri
           and John Shepherd},
journal  = {Distributed and Parallel Databases:
           An International Journal},
year     = {2004},
volume   = {15},
number   = {3},
pages    = {237--275},  }
```

which produces:

Anne Ngu, Banchong Harangsri, and John Shepherd. Query size estimation for joins using systematic sampling. Distributed and Parallel Databases: An International Journal, 15(3):237-275, 2004.

The Bibliography

Example: BibTeX and reference for a web page:



```
@misc{ postgresql,
author = "PostgreSQL Global Development Group",
title  = "PostgreSQL: The world's most advanced
          open source database"
howpublished = "http://www.postgresql.org/",
note   = "Accessed: 7 March 2008", }
```

which produces:

PostgreSQL Global Development Group. PostgreSQL: The world's most advanced open source database. http://www.postgresql.org/. Accessed: 7 March 2008.

Last Words

- Use Thesis Course Web Site
- Start as soon as possible
- Contact your supervisor ASAP
- Successful Thesis A —> successful Thesis B
- Be positive and enthusiastic. It’s your thesis.

Abstract

Finding the likeliest assignment to variables is a problem encountered in many areas. These include bioinformatics, natural language processing and computer vision. Mathematically, this is Maximum a Posteriori (MAP) inference. One way to solve this problem is to use probabilistic graphical models (PGMs). They encode complex joint probability distributions over a large number of variables in a compact form. One ongoing challenge is to develop efficient algorithms for performing inference on these models.

There exist software libraries that implement such algorithms, but many use a dense representation of the data. Specifically, tables are used for subsets of variables that all have statistical dependencies on each other. These permute all the assignments to the variables and associate with each assignment a cost.

Some libraries, for specific applications (e.g. computer vision), have specialised code for when the tables are sparse (many of the costs in a table are the same). Rather than write specific code for each application, this thesis presents a novel way to take advantage of sparsity with the only assumption being that the problem is sparse.

This thesis discusses the relationship between the amount of memory used to represent a discrete conditional Markov random field (CRF), a type of PGM, and the resulting time needed to perform particular types of inference. Using coordinate