Finding and Reviewing Literature

Will Smith

October 2013

Surveying Literature. Why bother?

1. Understand your research field and how to work in it

You need to show general knowledge and specific expertise

2. Don't reinvent the wheel

Your PhD contribution must be original

3. Don't reinvent the square wheel

It may already have been tried and failed

4. You aren't alone

Networking is important and exploitable

5. Build effectively on others' work

Contribute to a community and a body of work

Why bother?

Justification of originality

- identify a gap in the (academic) market for your topic
- you must show that your research hypothesis is original

Student A

I asked the guys in my office if they thought someone had done X. They didn't know of anyone who had, and since my dog didn't know either, I thought I'd have a bash.

Better

Despite a wide-ranging search of the available literature (reported in this chapter), I could find none that addresses issue X. This thesis investigates X.

Why bother?

To identify opposing views

- Get arguments for and against

Know where your work fits in

- I am part of the X community and my contribution is Y

To identify theory and methods you may need

- unifying frameworks, experimental design

Ideas on conducting literature reviews

`www.deakin.edu.au/library/findout/research/litrev.php`

What sources of information are there?

Primary sources

- journals, conference & workshop proceedings, doctoral & masters theses, technical reports, standards & codes of practice

Secondary sources

- journal review papers, thesis review chapters, abstracts, abstract reviews, monographs, ACM & IEEE digital libraries

Tertiary sources

- bibliographies, on-line bibliographies, subject indexes

Indicators of quality

ACM & IEEE imprimatur

- ACM Computing Surveys
- IEEE Trans. on Pattern Analysis and Machine Intelligence
- IEEE Real-Time Systems Conference

Impact factors

- `www.cs.york.ac.uk/rae/impact05.htm`

Citation counts

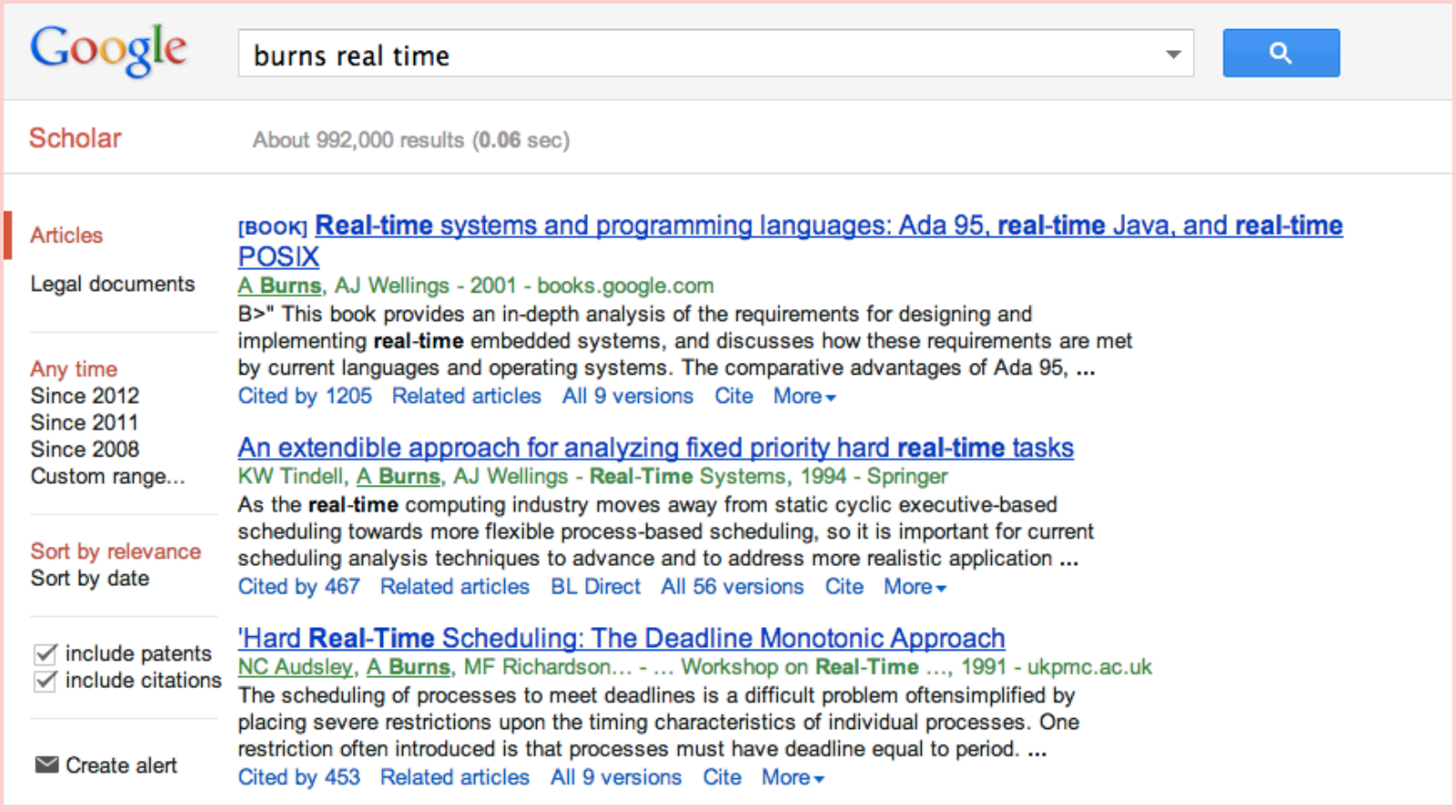
- `scholar.google.com/`
- ISI Web of Science

What else?

- Where do the majors players publish? Academic opinion? Acceptance rates?

Example: Google Scholar

Google "`Burns + real + time`"



Example: Google Scholar

What to do next

- follow citations (both directions)
- *read* the papers and books
- check out survey papers and literature chapters!

High profile general outlets

Top-end science

- most scientists, and even some members of the public, know
  - **Nature**, **Science**
- very significant scientific results
- not much computer science, but some
- most scientists would almost kill for these publications

Quality Journals and Conferences

- **Vision/pattern recognition**
  - IEEE Transactions on Pattern Analysis and Machine Intelligence
  - IEEE International Conference on Computer Vision, European Conference on Computer Vision, IEEE Computer Vision and Pattern Recognition
- **Neural Networks**
  - IEEE Trans. on Neural Networks, Neural Networks Journal
  - International Conference on Artificial Neural Networks, Neural Information Processing
- **Robotics**
  - IEEE International Conference on Robotics and Automation
  - IEEE/RSJ International Conference on Intelligent Robotics and Systems

Quality Journals and Conferences

- **Real-time systems**
  - Real-Time Systems Journal
  - IEEE Real-Time Systems Symposium
  - European Conference on Real-Time Systems (ECRTS)
- **Quantum computation and quantum information**
  - Physical Review Letters, sometimes Science or Nature
  - Quantum Information & Computation
- **Artificial intelligence and machine learning**
  - Journal of Machine Learning Research, Journal of Artificial Intelligence Research
  - International Conference on Machine Learning
- **Agents**
  - Autonomous Agents and Multi-Agent Systems
  - International Joint Conference on Autonomous Agents and Multi-Agent Systems

Quality Journals and Conferences

- **Formal Methods**
  - Formal Aspects of Computing
  - Formal Methods Symposium
- **Computational linguistics**
  - Journal of Computational Linguistics
  - International Joint Conference in AI (IJCAI)
- **Functional Programming**
  - Functional Programming Languages and Systems
  - Journal of Functional Programming
  - ACM International Conference on Functional Programming
- **Human-Computer Interaction**
  - International Journal of Human-Computer Studies
  - ACM Chi
- **Theoretical Computer Science**
  - Journal of the ACM
  - Journal of Computer and System Sciences
  - Foundations of Software Science and Computation Structures

Some subjects are just huge

Umbrella publication outlets

- IEEE Transactions on Software Engineering
- Journal of Systems and Software
- IEEE International Conference on Software Engineering

Specialist sub-disciplines

- IEEE International Symposium on Requirements Engineering
- Software Testing Verification and Reliability
- Formal Methods in System Design
- Formal Methods Symposium
- Information and Computation
- International Conference on Computer-Aided Verification

Some subjects are just huge

Umbrella subjects, e.g., critical systems

- software, hardware, human computer interaction, protocols. . .
- research will generally be in some sub-domain, but umbrella conferences and journals are potential outlets as well as domain-specific ones
  - IEEE Transactions on Critical Systems
  - Reliability Engineering and System Safety

How do you choose?

- your supervisor will know where best to publish

Exceptions

Be critical (both positive and negative)

- fine papers can appear in strange places
- don't assume a paper in a good journal/conference is authoritative
- citation culture varies between disciplines
- **be sceptical**
- example: Fischer
  - classification of state-rich behavioural formal methods
  - deeply flawed
  - widely cited

What are workshops for?

- to promote discussion
- get a topic off the ground
- cater to sectional interest
- can be very high quality
  - Dagstuhl
- can be of dubious quality

Theses

- you're not the first
- others will have reviewed the literature
- thesis lists
  - British Reports, Translations and Theses Bulletin
  - York theses are available on-line
  - most universities do the same
- but don't assume that they found all of the relevant literature

Research Reports

- quality varies
- most departments have internal Technical Report series
- typically reviewed before publication
- purpose
  - record position statements
  - record work in progress
  - record work difficult to publish in full
    - proofs, experimental results
  - scientific record
  - useful unusual work

Codes of Practice and Standards

- important sources in some domains
  - DO-178B, DEF STAN 00–55, ...
- some on-line
  - UK defence standards
  - `www.dstan.mod.uk/`
- some available in library
- some available from standards organisations
  - British Standards Institute
  - International Standards Organisation
  - Some cost!

What on-line sources are there?

`liinwww.ira.uka.de/bibliography/`

- artificial intelligence
- compiler technology, programming languages, type theory
- database research
- distributed systems, networking and telecommunications
- computer graphics and vision
- logic programming
- (computational) mathematics
- neural networks
- object-oriented programming and systems
- operating systems and parallel processing
- software engineering and formal methods
- theory/foundations of computer science
- typesetting
- miscellaneous
- technical reports

Bibliographies

- Google scholar
- DBLP

Search DBLP for "William A. P. Smith"

William A. P. Smith

List of publications from the DBLP Bibliography Server - FAQ
Other views: by type - by year (modern) - classic-C

Ask others: ACM DL/Guide - CSB - MetaPress - Google - Bing - Yahoo

	2012
j9	Ankur Patel, William A. P. Smith: Driving 3D morphable models using shading cues. Pattern Recognition 45(5): 1993-2004 (2012)
j8	Ankur Patel, William A. P. Smith: Automated Construction of Low-Resolution, Texture-Mapped, Class-Optimal Meshes. IEEE Trans. Vis. Comput. Graph. 18(3): 434-446 (2012)
c61	Oswald Aldrian, William A. P. Smith: Inverse Rendering of Faces on a Cloudy Day. ECCV (3) 2012: 201-214
c60	Oswald Aldrian, William A. P. Smith: Model-Based Ambient Occlusion for Inverse Rendering. ICIAR (1) 2012: 338-347
	2011
j7	Jing Wu, William A. P. Smith, Edwin R. Hancock: Gender discriminating models from facial surface normals. Pattern Recognition 44(12): 2871-2886 (2011)
c59	Fang Liu, Edwin R. Hancock, William A. P. Smith: Facial Expression Recognition Using Nonrigid Motion Parameters and Shape-from-Shading. CAIP (2) 2011: 105-113
c58	Ankur Patel, William A. P. Smith: Simplification of 3D morphable models. ICCV 2011: 271-278

Pre-print Sites (I)

- particular domains often have pre-print archive
  - cryptology ePrint Archive
  - `eprint.iacr.org/`
- papers generally submitted for publication elsewhere
  - comment before submission to conference/journal
- only lightly refereed
  - claim to first publication
- e-Print archive mirror
  - `uk.arxiv.org/`
  - good for quantum computing and information processing
  - weak for traditional computer science
- other subject specific sites

Research Group Web-sites

- find leading research groups
- publications available on-line
  - copyright restrictions
  - on-line delay
  - on-line first
- some researchers ignore copyright restrictions
- some excellent sites
- some very poor sites
- good local example
  - **RTSG:** `www.cs.york.ac.uk/rts/`

Department Web-site

- we put copies of theses and yellow reports on the ftp server
  - `www.cs.york.ac.uk/ftpserver.html`
  - (access via http or ftp)

Concerns about on-line sources

- if it's not on the web today, it doesn't exist
- if it's on the web today, it may not be tomorrow
- no refereeing for most sites

Issues in conducting a literature survey

- how do you assess a paper?
- criticism is not about being nasty
- balanced and reasoned assessments about contributions

Student B

The annotation language used in this paper was subsequently greatly improved by X [1] (virtually all modern work uses X's annotation language), but the fundamental approach remains unchanged to this day, e.g., [2,3,4,5,6,7,8].

Better

X maintained that, although he could not prove it, VIC was almost certainly true [1]. Many others agreed, expressing hope that it would soon be proved [2,3,4,5,6]. Only Y expressed doubt, commenting that mathematical fact isn't subservient to hope [9]. This proved well-founded, when in 2004, Z produced a counter-example [10].

Criticising Work

- Make sure you understand the paper and its context

Student C

I can't begin to indicate the naïvety of X's compiler work [1]. The optimisation is nonexistent, it makes no attempt whatsoever to address the important issues of concurrency (how bad is that?), and the error handling is atrocious. I could have done better myself.

Better

In 1965, X produced the first compiler for language Y [1]. Although by modern standards Y is not a particularly complicated language, it nevertheless posed several new problems at the time, such as Z. Some of the solutions adopted by X feature in most modern compilers. X commented that interpreting error messages required great skill, and that error handling would . . .

Criticising Work

- how influential has the paper been?
- you may have to judge after further reading
- how well reasoned is the argument?
- does the paper adequately refer to earlier work?
  - people have often discovered similar results or done similar work independently
  - X may simply not have known about Y at the time
    - X doesn't refer to the work of Y
    - X chooses not to acknowledge the work of Y

Criticising Work

- issues of scope:
  - identify limitations
  - identify assumptions (explicit or implicit)
- freedom from emotive language?
- presentation, structure, organisation
- how fair are the author's criticisms?
- who refers to the paper?
- is the research method sound?
- what is the research method??

Criticising Work: Methodology

- how convincing is the work?
- proof is a good way to convince
  - are the proofs elegant?
  - new results are good
  - new proofs may be good too
  - some fields are not keen on this

Criticising Work: Methodology

- how convincing is the work?
- do the experiments provide adequate evidence?
- don't expect papers to provide absolute answers to questions

Student D

The above results were obtained using a network of 10 PCs running for six months. A rigorous answer would require huge computational resources. The results presented, though not conclusive, provide evidence that further very large-scale experimentation has hopes of successfully resolving the issue.

Criticising Work: Methodology

Student E

I ran the algorithm 5 times on the iris dataset [1]. It produced an excellent answer every time, far quicker than other techniques [2,3]. I conclude that this is definitely the way to go. Everyone should use my new algorithm.

- what about other datasets?
- what about size of data sets?
- training versus evaluation data?
- statistical significance of results?
- were [2,3] intended for this type of problem?
- is the algorithm tuned and evaluated against the vanilla variants of [2,3] (the most basic versions)?

Telling a Story

- 10 weeks reading for your literature review seminar
- tell a story for the informed reader
- you must provide
  - introduction to problem area and motivation for your work
  - critical survey of the literature you have read
    - indicating seminal, highly influential papers
    - indicating historical development of the field
    - highlighting general agreements and disagreements, adequacy of assumptions, limitations, etc
    - pointers to unresolved issues in the field
    - adequate references

Breadth versus depth

make decisions on breadth and depth of survey

security
security protocols

- *types of security protocols*
- *verification techniques for security protocols*
  - *process algebras and model checking*
  - *use of logics*
    - *logics of knowledge*
    - *logics of belief: BAN logic, GNY and SVO*

Your Reference Index

- keep citation details and notes
- many sites have BibTeX entries for papers

```
Title of paper
Journal name
Publisher
Volume
Number
Page numbers
Month
Year
```

- Add references as you write (not add later)
- For urls – date of last access

And finally. . .

- facilities available in library (and leaflets on what is available)
  - reference databases, on-line journal information, catalogues of available journals, using the British Lending Library etc
- This seminar intended as short consciousness raising.
- further information on literatures searches on the web
  - e.g., Critically Analysing Information Sources at
    - `www.library.cornell.edu/olinuris/ref/research/skill26.html`
  - The Management of a Student Research Project. Sharp, Peters and Howard. Open University. Gower.

Thanks

This lecture has been given by John Clark, Jim Woodcock and Alan Burns who have all contributed to the slides