

---

# CS4040: Research literature

*Ehud Reiter*

# CS4040 Projects

---

- Projects MUST include experiments
  - » Where you gather empirical data
  - » Eg, time taken to do something
  - » Accuracy of search results
  - » Etc
- Cannot just be lit review or policy analysis

# Scientists build on prev work!

---

- Isaac Newton: “If I have seen further it is by standing on the shoulders of giants”
- Must know prev work to build on it!
- Also check that your idea has not been done by someone else
- Must understand scientific literature

# Literature is huge

---

- Natural Language Processing – main archive contains over 100K papers
  - » No one can read all of it
- Must find relevant papers

# Content

---

- **Literature review**
- Types of research literature
- Finding related work
- Reading related work
- Citing related work

# Literature Review

---

- Survey of research literature within the subject or topic your are investigating
  - » Synthesise/summarise key papers
  - » Critically analyse the information
  - » Identify gaps
  - » Understand limitations
- Present above in an organised fashion

# Related Work

---

- Your project report must describe related work
  - » Both dissertation and CS4040 project
- You must find and summarise the most relevant existing research on your topic
  - » Dissertation: You must explain how your work is different from previous work

# What to Read

---

- Depends on knowledge area
  - » Ask supervisor
  - » Start broad, eg textbooks
  - » Key (eg highly cited) papers in area
  - » Interesting references
- Read throughout the project
  - » Not just beginning or end



# Content

---

- Literature review
- **Types of research literature**
- Finding related work
- Reading related work
- Citing related work

# Many types of research lit

---

- Workshops and Conference paper
- Journal articles
- Books
- Theses
- Patents
- Web pages, including Wikipedia
- Preprints (Arxiv)
- Commercial state of art

# Peer Reviewed

---

- Most academic papers are peer reviewed
  - » Experts in field assess whether it is correct and worth publishing
  - » Quality assurance
- Some stuff is not peer-reviewed
  - » Read with caution...
  - » Could be real, could be marketing hype

# Citations

---

- Can assess paper qual/impact by how often cited by other research papers
  - » Use Google Scholar
    - <https://scholar.google.co.uk/>
  - » 100+ citations: influential research
  - » 10-100 citations: respected research
  - » 1-10 citations: hum...
  - » 0 citations: look elsewhere
- Different rules for just-published papers

# Availability

---

- Most CS research papers are available online for free
  - » Use Google or Google Scholar to find
- Some papers available for free online via uni library
  - » Library catalogue search
- A few cost money (especially books)

# Workshop and Conference

---

- Researchers gather together at workshops and conferences, to present their ideas and findings.
- In Computer Science, these events usually have “proceedings” of peer-reviewed papers by attendees
  - » Taken seriously in CS
  - » Not in other fields such as Medicine

# Workshop and Conference

---

- Usually 4-12 page papers
  - » Short but dense
  - » Expect readers know about the field
- Usually timely
  - » Presents research within past year
- Conferences usually peer-reviewed
  - » Some workshops are, some are not
- Example: assess 1 paper

# Journal articles

---

- Longer articles which present major research findings
  - » Typically 10-25 pages
  - » Usually include intro material, so easier for a newcomer to understand
- Main publication route for most sciences



# Journal articles

---

- Quality mixed
  - » Many excellent journals
  - » Some dubious ones which make money by charging authors to print papers
  - » Your supervisor should know
- Reputable journals will show their “impact factor”
  - » <https://www.mitpressjournals.org/loi/coli>
  - » Be wary if not shown, or less than 0.5

# Books and Theses

---

- Several kinds
  - » Textbooks – explain a topic
  - » Monographs – detailed presentation of authors research
  - » Edited collection – collection of chapters by different authors, all on related topics
  - » PhD theses - Aber Uni theses in library
- May cost money...

# Books

---

- Most books are not peer-reviewed, but there are exceptions
  - » Your supervisor should know
- Books are not widely used in CS to present research findings
  - » Different in other fields, esp humanities

# Web Pages

---

- Wikipedia is a respected source, and is edited and reviewed to some extent
  - » CS articles generally pretty good
  - » Sometimes contains marketing
- Other web sites can be variable quality
  - » Just use ones you can trust

# Patents

---

- Patents are publicly available
  - » Easiest access via Google Patents
  - » Can be hard to read (legalese)
    - <https://patents.google.com/patent/US9135244B2/en>
  - » I look at the figures first
- Patents are formally reviewed by official patent examiners

# Preprints/ Arxiv

---

- Preprint servers allow anyone to upload paper into an archive
  - » No peer review
- Arxiv (<https://arxiv.org/>) popular in CS
  - » Some great papers, some of which are available anywhere else
  - » Some garbage papers
  - » I look for authors I trust, because they have great peer-reviewed papers

# Commercial State of Art

---

- Commercial products which represent state of the art
- Usually cited via URL (webpage)
- Not peer reviewed
  - » Sometimes marketing hype!!
- Can be useful
  - » Find useful technical material!

# Content

---

- Literature review
- Types of research literature
- **Finding related work**
- Reading related work
- Citing related work



# What do I read?

---

- 100K papers in NLP
  - » Similar for other fields
- Which 10-20 do I read?

# Ask!

---

- Your supervisor will probably know of important related work
  - » Also other lecturers, postdocs, PhD students
- Ask!!

# Review articles

---

- Review articles summarise key research in a field
  - » Excellent source of pointers to the research literature
  - » Wikipedia articles often list review articles (under references or further reading)
  - » Some reviews are better than others...
- Textbooks also list key research papers

# Search engines

---

- Google is often useful
- Google Scholar is a search engine specifically for research papers
  - » <https://scholar.google.co.uk/>
  - » Related articles feature
- Semantic Scholar also useful:  
<https://www.semanticscholar.org/>
- Many others

# Generative AI

---

- ChatGPT (etc) can suggest related work
- You must read these papers!!
- You **will fail** if
  - » You cite hallucinated paper
  - » You make hallucinated claims about what a paper says
  - » Both above common with GPT

# Citations

---

- If you have found an old paper, look for more recent papers that cite this paper
  - » Google Scholar
- Also look for more recent papers by same authors
  - » Follow them on X/Twitter???
  - » I do this

# Forums

---

- Forums such as Stackoverflow can be useful
  - » Especially for more practical stuff
- Depends on who responds...

# Library

---

- Aberdeen University library has access to a lot of content that must be paid for
  - » Books
  - » Some journals
- CS researchers usually access online



# Content

---

- Literature review
- Types of research literature
- Finding related work
- **Reading related work**
- Citing related work

# Research Papers

---

- Some papers are excellent
- Some papers have good content but are hard to read
- Some papers are correct but not very interesting
- Some papers are wrong

# Step 1: Read abstract

---

- Read abstract first
  - » Maybe introduction and conclusion?
- Decide if paper is relevant, interesting, and believable

# Step 2: Skim paper

---

- Skim through the paper relatively quickly
  - » Can you understand it?
  - » Is it trustworthy?
    - Sensible statistics? Cite papers by other researchers?
  - » Is it relevant and interesting enough to be worth reading in detail?

# Step 3: Read Carefully

---

- Focus on relationship to your work
- Write notes as you read
  - » Will help you write related work section of your report
  - » Contribution of paper
  - » Strengths and insights of paper
  - » Weaknesses and concerns about paper

# Step 4: Contact Author

---

- If you have unanswered questions, consider emailing the author
  - » Most researchers are very happy to discuss papers with interested students
  - » Keep in mind that many people will not respond immediately.

# Reading Papers is a Skill

---

- You get better with practice and experience
- Don't be afraid to ask for help

# Caveat: Some Papers Wrong

---

- A lot of dubious papers get published
  - » Replication crisis
    - [https://en.wikipedia.org/wiki/Replication\\_crisis](https://en.wikipedia.org/wiki/Replication_crisis)
  - » Some deliberate fraud
  - » Huge problem in science
- Unfortunately hard for even experienced researchers to detect...
  - » [https://doi.org/10.1162/coli\\_a\\_00508](https://doi.org/10.1162/coli_a_00508)



# Content

---

- Literature review
- Types of research literature
- Finding related work
- Reading related work
- **Citing related work**

# Proper Citation

---

- Cite paper in text

Reiter [2018] showed that...

- Include cited papers in reference list

## **References**

E Reiter (2018). A Structured Review of the Validity of BLEU. *Computational Linguistics...*

# Many different styles

---

- Depends on field and venue
  - » Sometimes numeric citations ([10]) instead of author-date (Reiter 2018)
  - » Many formats for reference list
    - (MLA) Reiter, Ehud. "A structured review of the validity of BLEU." *Computational Linguistics* 44.3 (2018): 393-401.
    - (APA) Reiter, E. (2018). A structured review of the validity of BLEU. *Computational Linguistics*, 44(3), 393-401.
    - (Vancouver) Reiter E. A structured review of the validity of BLEU. *Computational Linguistics*. 2018 Sep 1;44(3):393-401.
    - etc

# Use Latex

---

- Latex includes “bibtex” subsystem for citations and references
  - » Separate “bib” file which gives information about possibly cited papers
    - Each paper has a key
  - » “cite” markups in main file
    - Using the “bib” key

# Bib file

---

- Contains entries such as the following

```
@article{reiter2018bleu,  
  title={A Structured Review of the Validity of BLEU},  
  author={Reiter, Ehud},  
  journal={Computational Linguistics},  
  year={2018},  
}
```

- Reiter2018bleu is the key
- Title, author, journal, year: info about paper

# Bib file entries

---

- Can download from Google Scholar and other archives
  - » <https://dblp.uni-trier.de/>
  - » <https://aclanthology.info/> (for NLP)
  - » Etc

# Writing bib file entries

---

- Can write yourself if necessary
- Needed for web page reference

```
@misc{p2pwiki,  
title={Automatic Summarization},  
author={Wikipedia},  
howpublished={\url{https://en.wikipedia.org/wiki/A  
utomatic_summarization}},  
note={Accessed: 14 Sept, 2016}  
}
```

# Citing papers in Latex

---

- Need to include some stuff in your main latex file
  - » `\usepackage{natbib}`
  - » `\bibliography{bib file name}`
  - » `\bibliographystyle{style}`
    - `\bibliographystyle{named}` most common in CS
- Overleaf/latex tutorial



# Cite

---

- Two formats
  - » `\citet{reiter2018bleu}` - Reiter (2018)  
Reiter (2018) showed that...
  - » `\citep{reiter2018bleu}` - (Reiter, 2018)  
BLEU is not always reliable (Reiter, 2018)
- Default `\cite` is usually `\citep`

# Cite

---

- Citet is “text” citation; use when directly claiming that a paper says something
- Citep as parenthetical reference; use to cite papers that back up a claim
- Dividing line is fuzzy!