



# HOW TO FIND PAPERS



BUT WHY READ PAPERS?

# WHY?

- X Stand on the shoulders of giants
- X Don't reinvent the wheel
- X Do reinvent the wheel, but pro

David J Bland  
@davidjblnd

Follow

The next time you hear "Don't reinvent the wheel"



6:50 PM - 19 Aug 2015

2,861 Retweets 1,965 Likes

76 2.9K 2.0K

3



# HOW TO FIND PAPERS



Okay, now that I am interested  
in field X, I want to read papers.  
Where do I begin?

A photograph of a sunset over a body of water. In the foreground, the dark silhouettes of two people are visible, facing away from the camera towards the horizon. The sky is filled with warm, orange and yellow hues from the setting sun, which is partially obscured by a dark, silhouetted landmass or large tree line on the right side.

read

research

## IN A NUTSHELL

- X Know where the top quality venues for the field of interest are
- X Know how to recognize the top quality venues
- X Know which search engines and online bibliographies to consult
- X Know what survey papers and Systematic Literature Reviews are
- X Know how to manage your references for continued research



# VENUES

“Where do babies papers come from?”

## CONFERENCE PROCEEDINGS VS. JOURNALS

- X Conference: annual events where researchers actually travel to and meet others – papers are presented (oral presentations or posters), and published in what's called a “proceeding”.
- X Journal: periodicals that always accept submissions (no deadlines) – accepted papers are published in a series of “books”.
- X We will discuss finer details between proceedings and journals, when we look at “what happens to your paper after submission” (scheduled on 21 March)

## TOP QUALITY VENUES

- X Each field has its premier venues – the most prestigious journals and conferences. Each comes with history of the field, so there is no set rule that can decide what is the best. But...
  - For journals, ACM and IEEE transactions with long running history are usually respectable and reliable.
  - The often ill-reputed SCI and SCIE tend to be noisy approximations when it comes to CS, but worth consulting.
  - For conferences, ACM and IEEE sponsored events are usually more reliable

## VENUE QUALITY

- X Does exist: there are journals and conferences you can relatively safely ignore. But...
- X Just because a paper is published in a good venue, it does not automatically mean that it contains great research.
- X And vice versa!
- X Consider venue quality to filter out definitely sub-standard material, but focusing on only the very top would be a mistake.

## VENUE SCOPE

- X For CS topics, journals are still categorised rather broadly.
  - A single ACM and IEEE transactions for CHI, SE, etc, despite many and diversified subfields.
- X Conferences are more fine-grained.
  - The top conferences tend to be broader, so that everyone in a field can gather.
  - There are usually multiple solid second-tier conferences that are more specialised (e.g., for testing, maintenance, etc in SE)



Okay, I get that there are venues  
with different quality levels.  
How am I supposed to know  
which is the good one?



## QUALITY INDICATORS

“If you have to explain them, they are not good”  
- A joke about quality of jokes



Well, this is hard to “teach” -  
you’re just supposed to know.

:P

## Now SERIOUSLY...

- X They cannot be sorted – there is no total order, only vague and sometimes subjective partial orders
- X There are often field-specific cultural subtleties that are hard to fully capture
- X Different fields have different approaches towards publications

## GOOD VENUES ARE

- X Where good individual papers are regularly published
- X Where good researchers meet (conferences) or send their work to (journals)
- X Where good researchers serve in the committees
- X Where you get good reviews
- X Once you get familiar with an area, you \*should\* naturally develop a “feel” for good venues
- X There are some quality indicators that can help you (with caveat)

## CITATION COUNT

- X For individual papers, this is simply the number of times it has been cited by other publications (including self citation)
- X Hard for individuals to keep track of, so we rely on various bibliographic databases to compute it
  - Scopus: citation database run by Elsevier (publisher)
  - Web of Science: originally run by Institute for Scientific Information (ISI), a subsidiary of Thompson Reuter, now owned by a private company called Clarivate Analytics
  - Google Scholar: run by the big G – known to be the most generous count (hence the popularity).

## IMPACT FACTOR

- X Impact Factor of a journal is the yearly average number of citations to articles published in that journal during the last 2 years
- X Nature's IF in 2014

$$IF_{2014} = \frac{Citations_{2013} + Citations_{2012}}{Publications_{2013} + Publications_{2012}} = \frac{29753 + 41924}{860 + 869} = 41.456$$

## IMPACT FACTOR

- X Huge variances between fields: CS journals are usually on the lower end
  - IEEE Transactions on Mobile Computing: IF 2016 = 3.822
  - ACM Transactions on Computer-Human Interaction: IF 2015 = 3.22
  - IEEE Transactions on Software Engineering: IF 2016 = 3.272

## H-INDEX / H5-INDEX

- X For an individual researcher, her/his h-index is the largest number  $h$  such that she/he has at least  $h$  publications all cited at least  $h$  times each.
  - Many (valid) criticisms, but still widely used as a performance metric
- X For a conference or journal, h5-index is the largest number  $h$  such that at least  $h$  publications from that conference/journal during the last 5 years are cited at least  $h$  times.

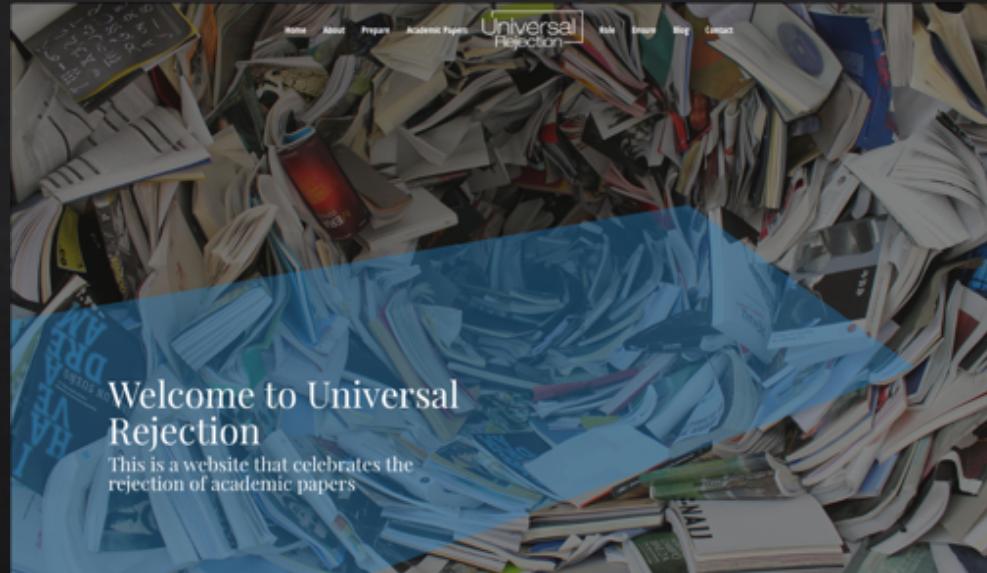
## H-INDEX FALLACIES

- X One of the most widely used bibliographic metric for individuals, BUT
- X It has a strong correlation with:
  - Experience: cannot compare junior and senior researchers
  - Community size: if more people are working in a field, it drives up the average h-index
- X It can be counter-intuitive:
  - If you work on a super-difficult stuff, there may not be that many people who would cite you

## ACCEPTANCE RATE

- X Acceptance rate is the ratio between published and submitted papers
- X Some perceive this as inversely correlated to the quality of conference
- X While this \*is\* used as a metric to evaluate conferences, don't read too much into acceptance rates
  - At the very top venues, acceptance is eventually noisy (we will discuss more when we talk about the pipeline)
  - Some actively argue that conferences should accept more so that ideas are freely exchanged

IF YOU TAKE THE “HARDER THEREFORE BETTER” MENTALITY  
TO ITS LOGICAL CONCLUSION...



<http://www.universalrejection.org>



# DON'T GAME AND BEWARE OF THOSE WHO GAME

- X Any metric can and will be exploited and \*gamed\* – by both individuals and venues
- X Beware those who do
- X Moderate your self-citations
- X You \*can\* recommend additional citations as a reviewer, for the good of the science, but do not be silly ->

 **Mamas Mamas**  
@mmamas1973 [Follow](#) 

I received a review back from a paper i submitted today from a cardiac journal with an IF>4. They asked us to cite 60 😳 papers that have been published in their journal before. Has anyone ever been asked to do this? do you think this is appropriate? and when? NO NAMES PLS.

Dear Dr. [REDACTED]

Reviewers have now commented on your paper. You will see that they are advising that you revise your manuscript. If you are prepared to undertake the work required, I would be pleased to reconsider my decision.

For your guidance, reviewers' comments are appended below.

If you decide to revise the work, please submit a list of changes or a rebuttal against each point which is being raised when you submit the revised manuscript.

Below you find a list of already published articles in [REDACTED] Please add these to your list of literature.

Your revision is due by 19-09-2018.

To submit a revision, go to [https://\[REDACTED\]](https://[REDACTED]) and log in as an Author. You will see a menu item call Submission Needing Revision. You will find your submission record there.

Yours sincerely  
[REDACTED]  
Editor-in-Chief  
[REDACTED]

8:58 PM - 20 Aug 2018

# RANKINGS

- X “All rankings are wrong, unless I’m at the top” :p
- X Do not trust them too much, but do consult them to know what is NOT near the top
- X Google Scholar maintains venue rankings based on h5 index:  
[https://scholar.google.co.uk/citations?view\\_op=top\\_venues&hl=en](https://scholar.google.co.uk/citations?view_op=top_venues&hl=en)
- X Microsoft compiles their own ranking:  
<https://academic.microsoft.com>
- X Computing Research & Education, run by Australian government, grades conferences: <http://portal.core.edu.au/conf-ranks/>



# SEARCH ENGINES

“If you type `google` into Google, you \*can\* break the Internet”  
– IT Crowd ([link](#))



## TYPES OF SEARCH ENGINES (RELEVANT TO CS)

- X First-party: run by academic associations (IEEE, ACM) or publishers themselves
  - ACM Digital Library (<https://dl.acm.org>)
  - IEEE Xplore Digital Library (<https://ieeexplore.ieee.org/Xplore/home.jsp>)
  - Elsevier ScienceDirect (<https://www.sciencedirect.com>)
  - Springer Link (<https://link.springer.com>)
- X Third-party: run by external organisations
  - Google Scholar (<https://scholar.google.com>)
  - ResearchGate (<https://www.researchgate.net>)
  - Allen Institute's Semantic Scholar (<https://www.semanticscholar.org>)



## BIBLIOGRAPHIES

- X Closer to index or catalogue than full database
- X Sometimes compiled by experts in the area
- X Digital Bibliography & Library Project (DBLP), owned by University of Trier, edited by Leibniz Centre for Informatics (<https://dblp.uni-trier.de>): perhaps the most authoritative index of individual researcher's publication record
- X Penn State University's CiteSeerX (<http://citeseerx.ist.psu.edu>)
- X Can be much more topic specific: when done properly by individuals, these can be great service to the community
  - o [Mutation Testing](#), [Genetic Programming](#), [Search Based Software Engineering](#), [Test Oracles](#), etc



## SEARCH STRATEGIES

### If you are after an individual:

Start with Google Scholar to see publication trends, etc

Specific papers may or may not be available on GS: move to ACM/IEEE/publisher sites if not

Consult DBLP for more accurate records

### If you are after a topic or a paper:

Start with the relevant publisher site (ACM/IEEE/publisher): publications are better linked – cites & cited by, as well as by event series

If fulltext is missing, you may want to use RG to request one; you can also write a polite email.

### If you are after a specific event:

DBLP or ACM/IEEE is your friend



## HOW TO KEEP UP

- X Some journals provide RSS feeds
- X You can follow either topic or people on Google Scholar
- X Some researchers blog about their, or even other people's, research actively
  - [The Morning Paper](#) by Adrian Coyler
- X Connect to people on SNS – but be SNS-savvy!
- X Long term, you also keep track of recent research by
  - Being invited to review journal papers
  - Being invited to Technical Program Committees



## SNOWBALLING

- X Start from a core set of papers on a topic
- X Follow multiple dimensions:
  - Papers from the same venue, both before and after what you already read
  - Papers from the same group of authors
  - Papers cited by what you already read, papers that cite what you already read
- X Eventually you will reach a reasonable closure

4.

## SURVEYS AND SLRs

“Survey says Americans getting tired of surveys”  
– Unidentified newspaper headline ([link](#))

## SURVEY PAPERS

- X Survey papers literally “survey” the current landscape of a research area: what are the challenges, what papers have been written to address the challenges, what is the future direction
- X They are NOT simple catalogues of existing papers
  - Bad surveys will simply list of papers out there
  - Good surveys will educate the reader with helpful perspectives
- X “Regular research papers are a description of your own research. A survey paper is a service to the scientific community” – Lotzi Böloni, University of Central Florida ([link](#))

## VALUES OF SURVEY PAPERS

- X For readers: instead of reading sometimes 100+ papers, you can read a digest and get a broader view on a topic
- X For writers: when done right, you will do a great service to the community, attract citations, and get yourself known
  - Your thesis (especially for PhD) should contain a review of literature – write it as if writing a survey (or, in fact, try to publish that part)

## SYSTEMATIC LITERATURE REVIEWS

- X SLRs are in-between regular research papers and surveys. They DO ask specific research questions, but instead of conducting own studies, they look at existing results in the literature (a meta-analysis, if you like)
- X For example: “No Amount Of Alcohol Is Good For Your Health, Global Study Says ([NRP](#))”
  - A meta-analysis using 694 data sources ([The Lancet](#))
- X Fewer precedences in CS, perhaps due to the extreme variability of software: but there are good examples



# REFERENCE MANAGEMENT

“What happened to the dust on my desk? I wrote some important phone numbers on it!”



## REFERENCE MANAGEMENT

- X Do not just \*consume\* papers by reading them: organise them into your own little database
- X There are many tools that help you with this
  - o Papers: <https://www.readcube.com/papers> (Mac, iOS), ReadCube: <https://www.readcube.com> (PC, Android, Web)
  - o Mendeley: <https://www.mendeley.com>, (all platforms) acquired by Elsevier
  - o BibDesk: <https://bibdesk.sourceforge.io> (Mac) DB based on BibTeX format
  - o JabRef: <http://www.jabref.org> (all platforms) DB based on BibTeX format
  - o EndNote: <https://endnote.com> (PC & Mac), easier integration with MS Word
- X If you can, make short notes of your impression



Your “reading” reference management tool should also serve you as your “writing” reference management tool.

## SUMMARIES

- X Try to develop your own “feel” for different venues
- X Understand quantitative metrics, but also know that they can be gamed
- X Do take into account existing rankings and reputation, but do not put blind faiths in them
- X Know which tools to use for different purposes
- X Manage what you read: they are ingredient for your writing in the future