

# Working with Academic Literature

Search, Screen, Read, Appraise

Jörg Cassens

Academic Literacy  
Winter term 2019/2020



## 1 Approach

### Phases & Theses

#### Work phase

Researching the topic

Exposition

Analysis

Implementation

Evaluation

Proof-editing

Presentation

#### Why do literature research?

Classification, clarification

Quick research, unfiltered literature (search engines, manuals, textbooks)

Systematic research

Researching current issues

Systematic research

Complete references

Reduction

Write everything down from the beginning and reference it. All scientific work needs a part “related work” that outlines the state of research & development.

### Type of Literature I

- Monographs
  - Textbooks
- Collections
  - Conference proceedings
  - Collections on specific topics
  - Encyclopedias
  - Special purpose books
- Journals
- News media
- Limited citability → “grey literature”
  - Websites
  - Technical Reports
  - Working Papers etc.
- Non citable
  - impossible to get access to for third parties (meeting minutes, lecture notes, personal communication)

## **Type of Literature II**

- Original article
  - Information based on original research
- Technical notes
  - describe a specific process, technology or method
- Survey
  - Detailed analysis of recent research on a specific topic
- Commentary
  - short article with author's personal opinions
- Editorial
  - often short review or critique of original articles
- Letter to the Editor
  - short & on subject of interest to readers

## **Relevance of Literature**

- Tendency
  - Scientific journal, peer-reviewed
  - Conference proceedings (usually reviewed in informatics)
  - Books and chapters (with or without review)
  - Reports usually without review
  - Working Papers usually without review
  - Popular science, usually without review
- But: (Almost) more important than the type of publication is the publication channel
  - Impact Factor
  - Acceptance Rate

A workshop on IJCAI (International Joint Conference on Artificial Intelligence), CHI (Annual CHI Conference on Human Factors in Computing Systems) or SIGGRAPH (Special Interest Group on Graphics Conference) is more important than a "Multi-conference in Las Vegas"

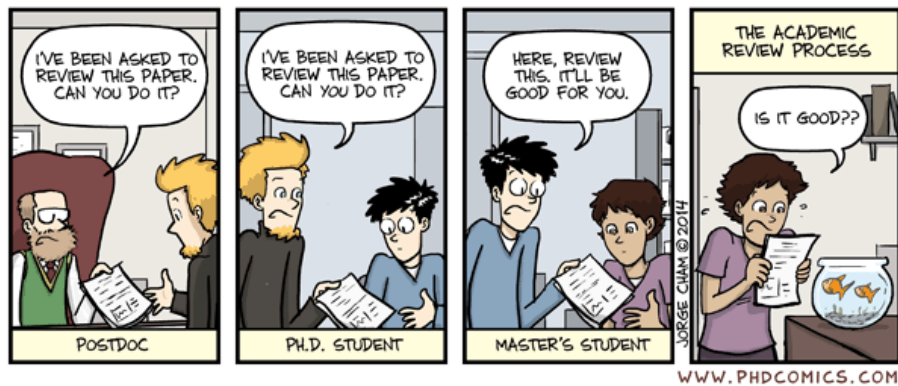
## **Review**

- Peer review: a publication's scientific value is assessed by other researchers in the area
- Potential factors:
  - Relevance of topics
  - Inter-subjectivity (Reproducibility and verifiability)
  - Free of contradictions
  - Understandability and readability
  - Method and innovative contribution
  - Context of existing work
  - etc.

## Versions

Usually, scientific work will go through certain phases: you might discuss first results in a workshop before presenting major results on conferences, thereafter giving an in-depth presentation of major results in journals before maybe participating in a book on the topic

Similar contents can therefore often be found in different media, the “most important” is the most citable (but it may be easier to access “less important” or earlier versions, e.g. on workshops)



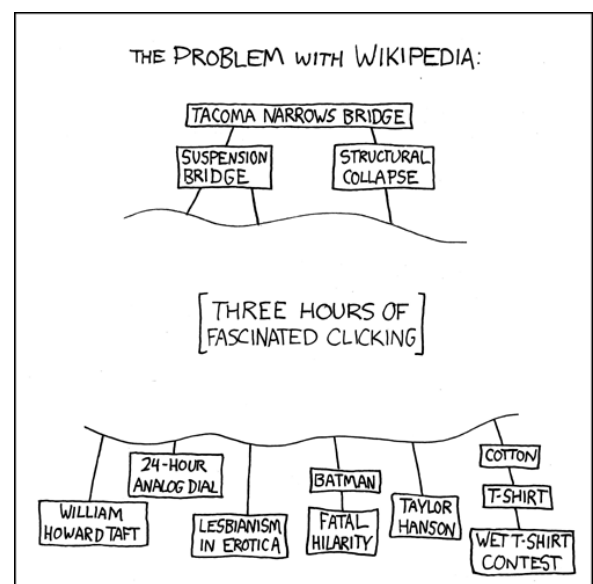
PhD Comics 1760

## Systematic Search I

- Goals:
  - What can help me with my problem?
    - \* Wheel 2.0: Do not reinvent the wheel
  - Has someone solved this problem before?
    - \* Demarcation
- Preparation:
  - What concepts and terms are important for my topic?
  - What are synonyms or related terms?
    - \* e.g. searching thesauri: Deutscher Wortschatz, WordNet
  - For German users: what are the English terms?
    - \* e.g. using LEO
    - \* partly Wikipedia
  - Relation between terms & concepts

## Systematic Search II

- Search
  - General search engines
  - Specialized search engines
  - Libraries
  - Bibliographies
  - (Internet-) book stores
- Acquire
  - (Digital) libraries
  - (Internet-) book stores
- Read
  - Skim & appraise



xkcd: the problem with wikipedia

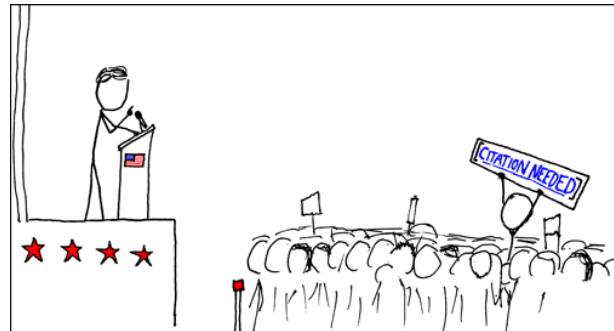
## 2 Search & Acquire

### Wikipedia

☞ **Dan Gillmor**

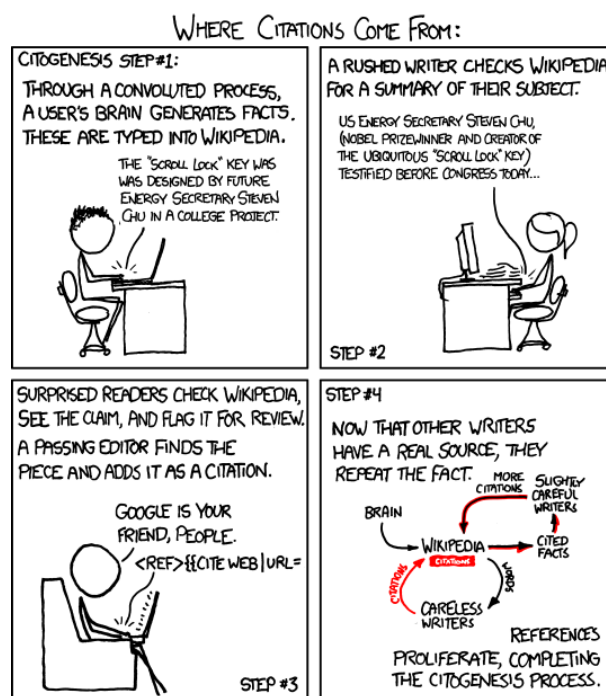
“Wikipedia is a great place to start researching and the worst place to stop”

- Use as starting point
- Not for deep research
- Not a primary source
- When to use?
  - First conceptualisation
  - Concept clarification
- No peer review
  - Read critically
- Look at the references
  - Use the source, Luke



☞ xkcd: Wikipedian Protester

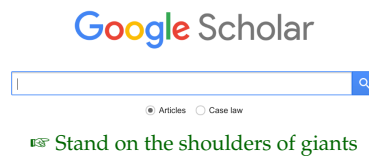
### Wikipedia



☞ xkcd: citogenesis

### Search Engines

- Google – known and loved by everyone
- Specialized version for academic literature: Google Scholar
  - ☞ [scholar.google.com](https://scholar.google.com)
  - Search for literature
  - With ranking (where published? cited how often?)
  - Links to sources for full texts
  - Author profiles
  - Includes lot of older “grey” literature like technical reports and makes them accessible



### Informatics Libraries: Publishers I

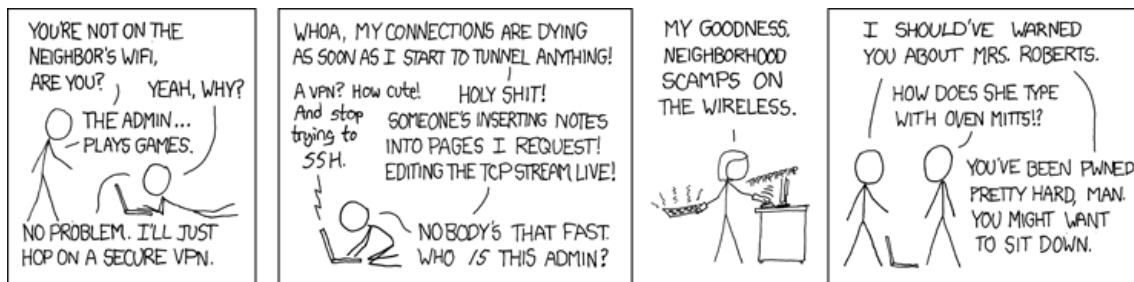
- Starting point Hildesheimer Katalog (HILKAT)
  - [hilkat.uni-hildesheim.de](http://hilkat.uni-hildesheim.de)
- ACM Digital Library
  - [dl.acm.org](http://dl.acm.org)
  - Full texts of all ACM Publications
  - Additional information, e.g. citations (BibTex, EndNote, ACMRef)
- Springer LINK Online Library in Computer Science
  - [link.springer.com](http://link.springer.com)
  - Over 100 journals in informatics
  - Contains LNCS (Lecture Notes in Computer Science) and sub-series (such as LNAI)

### Informatics Libraries: Publishers II

- IEEE Digital Library
  - [www.computer.org/csdl](http://www.computer.org/csdl)
  - all IEEE publications
  - Additional information (citations, e.g. EndNote)
- Wiley Online Library
  - [onlinelibrary.wiley.com](http://onlinelibrary.wiley.com)
  - Life sciences, medicine, social sciences & humanities
  - Additional information (citations)
- De Gruyter
  - [degruyter.com](http://degruyter.com)
  - Informatics, engineering, computational linguistics
  - Best supported by Unibibliothek
- Elsevier
  - [sciencedirect.com](http://sciencedirect.com)
  - Informatics, computational linguistics, medicine

### Note: VPN

- For those digital libraries where the university has access to full texts, access is restricted to the IP-range of the university
- To use them from outside, you need to use [VPN](#)
  - Make sure to configure that traffic to **all** IP addresses is routed via the university
  - You may have to start a new browser session



☞ xkcd: 1337 part 1



## Informatics Libraries: Independent I

- DBLP Computer Science
  - ☞ [dblp.uni-trier.de/db](http://dblp.uni-trier.de/db)
  - Uni Trier and Schloß Dagstuhl, Content CC0 (sic!)
  - More than 5800 conference and workshop series, more than 1600 journals in computer science
  - Almost 5 Mio publications with more than 2 million authors
    - \* Many links to author home pages
  - List of conferences and journals including tables of content
- The Collection of Computer Science Bibliographies
  - ☞ [liinwww.ira.uka.de/bibliography/](http://liinwww.ira.uka.de/bibliography/)
  - Karlsruhe Institute of Technology
  - More than 7 million references (2,3 GB BibTeX-entries)
  - More than 1 million of references contain URLs to online versions of articles
  - Comprised of several bibliographies, manually supported
  - Mostly journal articles, conference papers, technical reports
  - Partially includes full texts




## Informatics Libraries: Independent II

- Semantic Scholar
  - ☞ [semanticscholar.org](http://semanticscholar.org)
  - Allen Institute for Artificial Intelligence
  - Almost 180 million scientific papers from sources like PubMed, Nature, and ArXiv
  - AI analyzes research papers and pulls out authors, references, figures, and topics
  - Partially includes full texts
- Citeseer X
  - ☞ [citeseerx.ist.psu.edu](http://citeseerx.ist.psu.edu)
  - PennState University
  - Scientific literature digital library and search engine
  - Automated harvesting and processing of articles
    - \* author data
    - \* citation data
    - \* related documents
  - Partially includes full texts



## Researcher IDs

- ORCID
  -  [orcid.org](https://orcid.org)
  - Nonprofit organisation
  - Goal: “all who participate in research, scholarship and innovation are uniquely identified and connected to their contributions and affiliations, across disciplines, borders, and time”
  - Researchers can get a unique ID and curate their bibliographies
  - Links to other tools like DBLP, Scopus, ReasearcherID
- ResearcherID
  -  [researcherid.com](https://researcherid.com)
  - Clarivate Analytics (ex Thomson Reuters)
  - Each member is assigned a unique identifier
  - Self-curation of bibliographies
  - Integrates with Web of Science

## Citation Databases

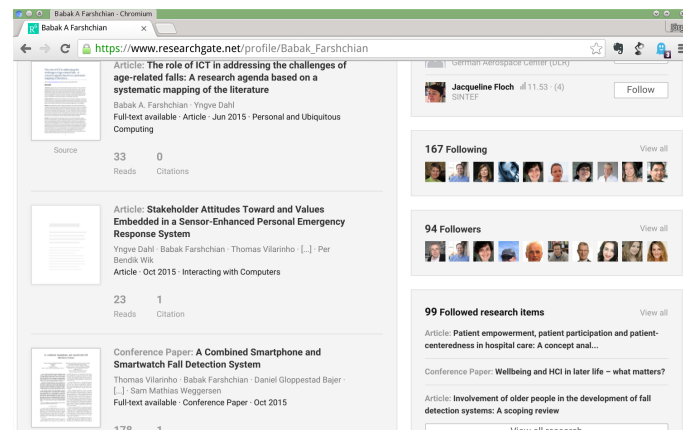
- CrossRef
  -  [crossref.org](https://crossref.org)
  - Publishers International Linking Association
  - Interlinks millions of items from a variety of content types, including journals, books, conference proceedings, working papers, technical reports, and data sets
  - Official Digital Object Identifier (DOI) Registration Agency
- Web of Science
  -  [webofknowledge.com](https://webofknowledge.com)
  - Clarivate Analytics (ex Thomson Reuters)
  - Comprehensive citation search
- Scopus
  -  [scopus.com](https://scopus.com)
  - Elsevier
  - Comprehensive citation search & ranking

## Open Access

- Open Access: free (as in beer) access to full texts
- [www.open-access.net](https://www.open-access.net)
- Open Access Journals sometimes called “Golden Road”
  - New economic model: author/institutions pays for publication, not access
  - Directory of Open Access Journals (DOAJ)
    - \*  [www.doaj.org](https://www.doaj.org)
- Self-archiving, often called “Green Road”
  - e.g. on institutional document servers
  - Sometimes includes peer-reviewing (Preprints, Postprints)
  - arXiv.org (Cornell University)
    - \*  [arxiv.org](https://arxiv.org)

## Social Networks

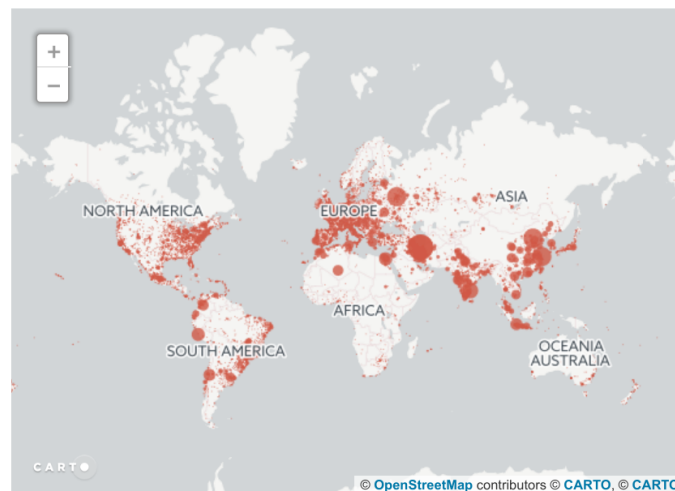
### Self-archiving on specialized social networks



ResearchGate, Academia

## Illegal Sources

They do exist.



Who's downloading pirated papers? Everyone. Science Magazine, 2016.



**Taylor R. Genovese** @trgenovese · Oct 23

I have some bad news for you. WILEY's CEO makes \$4 million. Elsevier's CEO? £10.5 million. Or \$13.8 million.

[telegraph.co.uk/business/2016/...](http://telegraph.co.uk/business/2016/...)

**M. Zeppel** @MelanieZeppel

Is this true? Does the #elsevier CEO really make \$4 mill p.a. ? While we write/review for free for them? Huh! @ElsevierNews  
[twitter.com/trgenovese/sta...](https://twitter.com/trgenovese/status/791111111111111111)



16



10



**Jörg Cassens** @cassens · Oct 23

As a reference for the numbers: [www1.salary.com/Mark-J-Allin-S...](http://www1.salary.com/Mark-J-Allin-S...)



**Taylor R. Genovese** @trgenovese

As long as the CEO of Wiley is making \$4,017,769/year off of academics work, I'll be posting anything I publish wherever the hell I please.

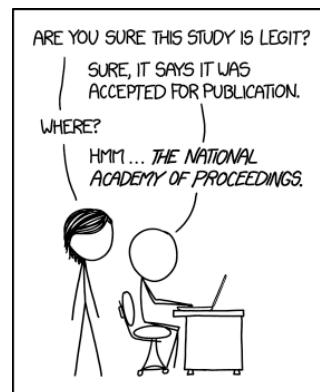
Taylor R. Genovese, Melanie Zeppel, Jörg Cassens

## 3 Read

### General Structure



- Title
- Abstract
- Introduction
- Methods
  - Analysis
  - Design
  - Implementation
- Results
  - Evaluation
- Discussion/Conclusions
  - Further Works
- Acknowledgements
- References

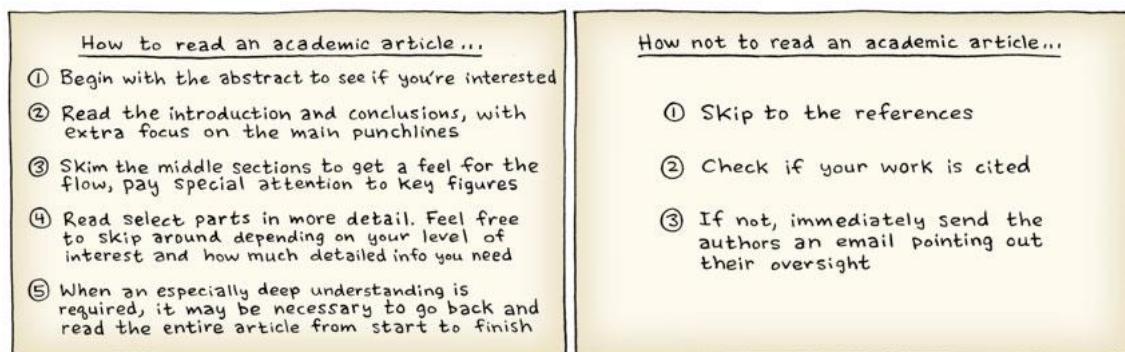


IF SOMETHING IS FORMATTED LIKE A SERIOUS SCIENTIFIC PAPER, IT CAN TAKE ME A WHILE TO REALIZE IT ISN'T ONE.

📖 xkcd: dubious study

## Skim & Read

- Do not read from start to finish
  1. Abstract
  2. Conclusions
  3. Introduction
  4. "the rest"
- Whether you continue with the conclusions or the introduction depends on your familiarity with the topics... and on the length of the sections



calamitiesofnature.com © 2012 Tony Piro

📖 Calamities of Nature

## Skim, Re-Read, Examine, Summarize

- Another approach championed by Natalia Rodriguez
- Four steps
  - Skim
    - \* Read for "big picture"
  - Re-read
    - \* Examine graphs, tables, figures
    - \* Interpret the data yourself
  - Examine
    - \* What problems are addressed?

- \* Why is it important?
- \* Is the method good?
- Summarize
  - \* Write a summary of key point in own words
- Elsevier Connect, [research4life.org](http://research4life.org)

### Three Pass Reading

- Read the Paper in three passes
  1. Quick scan to get a bird's-eye view of the paper to be able to answer the five C:
    - a) Category: What type of paper is this?
    - b) Context: Which other papers is it related to?
    - c) Correctness: Do the assumptions appear to be valid?
    - d) Contributions: What are the main contributions?
    - e) Clarity: Is the paper well written?
  2. Read the paper with greater care, but ignore details such as proofs
    - Afterwards, you be able to summarize the main thrust of the paper, with supporting evidence, to someone else
  3. Attempt to virtually re-implement the paper
    - Making the same assumptions as the authors, try to re-create the work
- S. Keshav: [How to Read a Paper](#)

### What to Look For

- Five elements to look out for:
  1. A significant question or claim
  2. A position in the academic debate
  3. An explanation of the research method or approach
  4. A presentation of the findings and argument
  5. A statement of the implications and contributions of the research study
- You should also aim to place a journal article within the broader academic debate
- Nicholas Carah and Eric Louw: [Guide to reading journal articles](#)

### Paper on Reading Papers

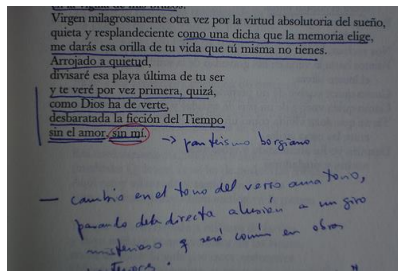
#### RV Subramanyam

"The reader should begin by reading the title, abstract and conclusions first. If a decision is made to read the entire article, the key elements of the article can be perused in a systematic manner effectively and efficiently. A cogent and organized method is presented to read articles published in scientific journals."

Subramanyam R V. Art of reading a journal article: Methodically and effectively. *J Oral Maxillofac Pathol [serial online]* 2013; 17:65-70. Available from: <http://www.jomfp.in/text.asp?2013/17/1/65/110733>

### Notes & Annotations

- Annotations
  - The paper advantage
  - Notes, marks, scribbles, Post-Its are sign of active, creative examination of the content
- Notes
  - Never read a text without pen & paper



cc by-nc Manuel Sanfuentes



cc by Kristian D.

## Excerpts

- Reproducing parts of a text
- Paraphrased or word-by-word
- WordNet synonyms: excerpt, excerpition, extract, selection

### 1. Orientation

- Get a grip of the structure of the text

### 2. Excerpt

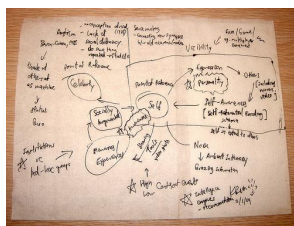
- Work with the text – “what is the contribution of this part of the text?” and “what are the main points made?”

### 3. Compact

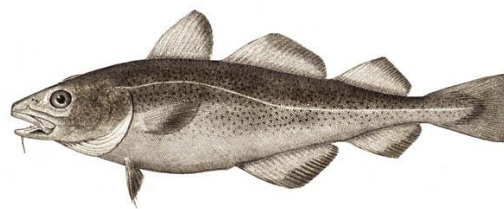
- Summarize the texts and the excerpts

## Visualisation & Focus

- Visualisation
  - Mindmaps, Concept Maps, ...
- Individual preferences
- Focus
  - You start with a limited amount of literature
  - You widen the search
  - But need to come back to your topic in the end



cc by-nc-sa Kevin Lim



cc 0

## Appraise

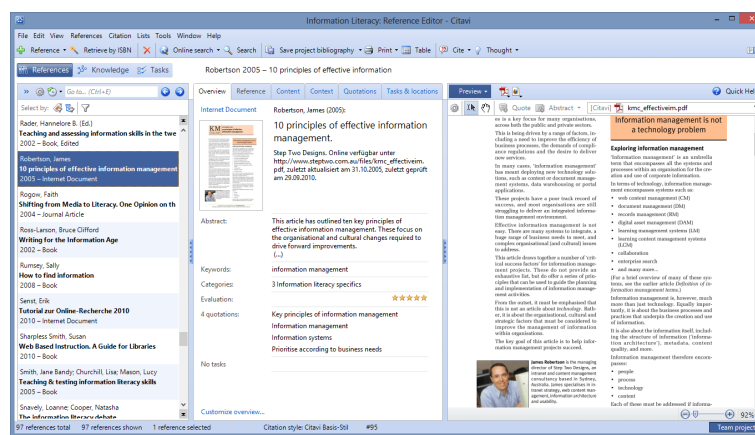
- There is no final and comprehensive key to judge and appraise a publication
- The “value” is always in the context
- A working paper without peer review can be very relevant for your own work when from a related area
- But most of you will not find the following article useful, even although published in a prestigious journal:
  - Bin Zou, Luoqing Li, and Zongben Xu: The generalization performance of ERM algorithm with strongly mixing observations. In: Machine Learning, Volume 75, Number 3/June, 2009.
- Citations and ranking might provide a clue, though

## 4 Manage

### Why Manage Literature?


- Write everything down from the start
  - what do you read?
  - what do you need to read?
  - what do you want to read?
- Make short notes & comments on what you read
- Make use of tools for literature management
  - Collecting references
  - Notebook and bookmarks
  - For the list of references
- Again: no “Silver Bullet”, different requirements

### Citavi

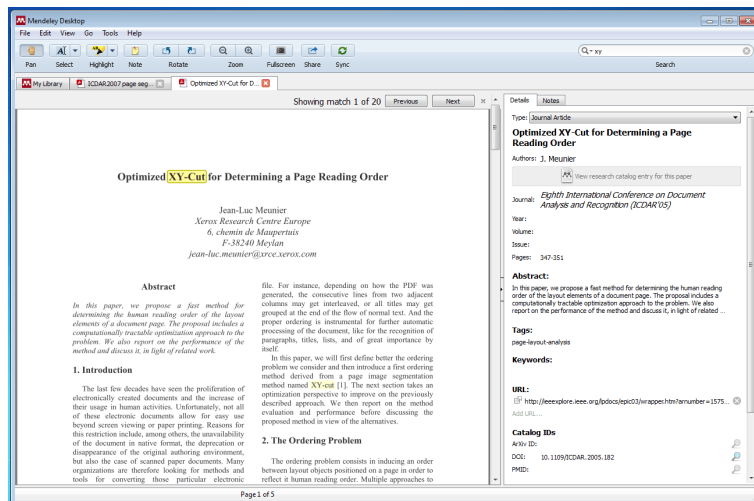


Source  citavi.com

### Citavi

- Commercial desktop software for Windows
- Campus license  [www.citavi.com/uni-hildesheim](http://www.citavi.com/uni-hildesheim)
- Manual or automatic import (ISBN, DOI), Citavi retrieves the bibliographic information
- “Citavi Pickers” for Internet Explorer, Firefox, Chrome, and Adobe Reader and Adobe Acrobat
- Find full text and use PDF in preview
  - Annotate, search
- Export to Word processors via RTF or use Microsoft Word plugin
- Works with LaTeX editors

## Mendeley

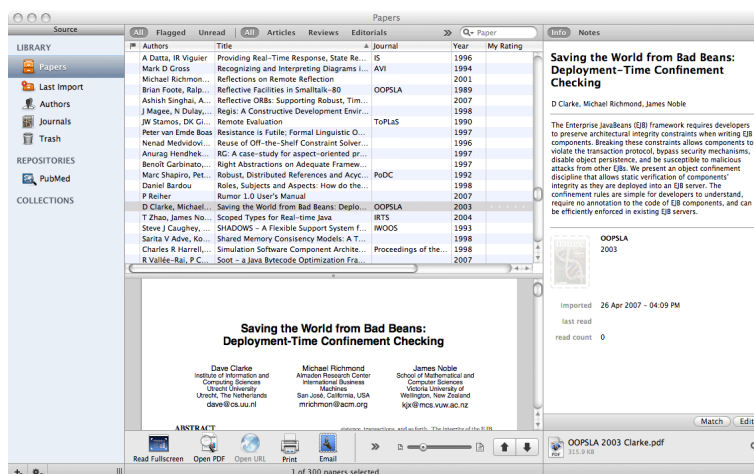


Source  mendeley.com

## Mendeley

- Commercial, free to use (Freemium)
- Owned by Elsevier
- Notes and Annotations for papers
- PDF preview
- Web, mobile (Android, iOS) or desktop (Windows, Linux, Mac)
- Plugins for Chrome, Firefox to add content from Web
- Citations and bibliographies in Microsoft Word, OpenOffice.org, and LibreOffice
- BibTeX export/file synchronisation
- Uses XMP Metadata in PDF
- Integrates with ORCID and others

## Zotero



Source  zotero.org

## Zotero

- Free and Open Source Software
- Online storage can be extended for a subscription fee
- Web or desktop (Windows, Linux, Mac)
  - Mobile support spotty (Android, iOS)
- Import and export citations from/to many formats, including BibTeX, BibLateX, RefWorks, MODS, COinS, Citation Style Language/JSON, refer/BibIX, RIS, TEI, several flavours of RDF, Evernote, and EndNote
- Export as formatted bibliographies in many formats
- Storing full text PDF supported
- Notes and Annotations for papers
- Collaboration in groups

## BibTeX & Jabref

The screenshot shows the JabRef application window titled 'JabRef - D:\TEMP\jabref-author.bib (BibTeX mode)'. The interface includes a menu bar (File, Edit, Search, Groups, View, BibTeX, Quality, Tools, Options, Help), a toolbar with various icons, and a search bar. Below the search bar is a table of bibliographic entries. The table has columns for #, entrytype, author/editor, title, year, journal/booktitle, bibteykey, and ranking. The entries are numbered 1 through 31 and include details such as the entry type (e.g., Article, Inproceedings), the author(s), the title of the work, the year of publication, the journal or book title, the BibTeX key, and a ranking of stars.

#	entrytype	author/editor	title	year	journal/booktitle	bibteykey	ranking
1	Article	Gambdo et al.	CORCIC II: A New Improved CORCIC Algorithm	2016	IEEE Trans. on Circuits and S.	GambdoKallisthromKurneEhNZ	
2	Inproceedings	Geiger et al.	On the Evolution of BPMN 2.0 Support and Implementation	2016	10th International IEEE Symp.	GeigerHamerLenthanEhNZ2016	★★★★☆
3	Inproceedings	Geiger et al.	Process Engine Benchmarking with Betsy - Current Status and Future Directions	2016	Proceedings of the 8th Centr.	GeigerHamerLenthanEhNZ2016	
4	Inproceedings	Nikol et al.	Service-oriented Multi-tenancy (SO-MT): Enabling Multi-tenancy for Existing Servic.	2016	10th International IEEE Symp.	NikolTraegerHamerEhNZ2016	
5	Phdthesis	Kopp	Partnerübergreifende Geschäftsprozesse und ihre Realisierung in BPML	2016		Kopp2016	
6	Article	Wiedinger et al.	Streamlined DevOps automation for Cloud applications using TOSCA as standa.	2016	Future Generation Comp. Syst.	WiedingerBrettenboescherKopp	★★★★☆
7	Inproceedings	Geiger et al.	BPMN Conformance in Open Source Engines	2015	Proceedings of the 9th IEEE I.	GeigerHamerLenthanEhNZ2015	
8	Inproceedings	Hamer et al.	Improving the Static Analysis Conformance of BPML Engines with BPMLint	2015	Proceedings of the 9th IEEE I.	HamerGeigerPreisingerEhNZ	★★★★☆
9	Inproceedings	Kopp et al.	A Domain-Specific Modeling Tool to Model Management Plans for Composite ap.	2015	Proceedings of the 7th Centr.	KoppBrettenboescherEhNZ	
10	Inproceedings	Brettenboescher et al.	A Modeling Concept to Integrate Declarative and Imperative Cloud Application Pr.	2015	CLOSER 2015 - Proceedings.	BrettenboescherEhNZKoppEhNZ	★★★★☆
11	Inproceedings	Brettenboescher et al.	A situation-aware workflow modelling extension	2015	Proceedings of the 17th Inter.	BrettenboescherHamerKoppEhNZ	
12	Inproceedings	Wagner et al.	Choreography-based Consolidation of Interacting Processes Having Activity-bas.	2015	CLOSER 2015 - Proceedings.	WagnerKoppLeymannEhNZ	
13	Inproceedings	Thummen et al.	Darstellung des Konzeptes - DMA Decentralized Market Agent - zur Bewältigung .	2015	45. Jahrestagung der Gesell.	ThummenKallisthromKurneEhNZ	
14	Inproceedings	Gustafsson and Johansson	Decimation filters for high-speed delta-sigma modulators with passband const.	2015	2015 IEEE International Sym.	GustafssonJohanssonEhNZ	★★★★☆
15	Inproceedings	Johansson and Gustafsson	Filter-bank based all-digital channelizers and aggregators for multi-standard v.	2015	2015 IEEE International Conf.	JohanssonGustafssonEhNZ	★★★★☆
16	Inproceedings	Alam and Gustafsson	Generalized division-free architecture and compact memory structure for resamp.	2015	European Conference on Cir.	AlamGustafssonEhNZ	
17	Inproceedings	Brettenboescher et al.	Hybrid TOSCA Provisioning Plans: Integrating Declarative and Imperative Cloud	2015	Cloud Computing and Seric.	BrettenboescherEhNZKoppEhNZ	
18	Inproceedings	Ingemansson and Gustafsson	On fixed-point implementation of symmetric matrix inversion	2015	European Conference on Cir.	IngemanssonGustafssonEhNZ	
19	Inproceedings	Johansson and Gustafsson	On frequency-domain implementation of digital FIR filters	2015	2015 IEEE International Conf.	JohanssonGustafssonEhNZ	★★★★☆
20	Article	Eshuis et al.	Service Outsourcing with Process Views	2015	IEEE Trans. Services Comput.	EshuisKortkampEhNZ2015	★★★☆☆
21	Inproceedings	Kopp et al.	Towards a Cloud-based Platform Architecture for a Decentralized Market Agent	2015	45. Jahrestagung der Gesell.	KoppKallisthromKurneEhNZ	
22	Inproceedings	Kopp et al.	Automated and Isolated Tests for Complex Middleware Products: The Case of B.	2014	Proceedings of the 7th IEEE I.	HamerRoelWitz2014	★★★★☆
23	Inproceedings	Hamer et al.	BPML Conformance in Open Source Engines: The Case of Static Analysis	2014	Proceedings of the 7th IEEE I.	HamerPreisingerWitz2014	
24	Inproceedings	Roel et al.	Performance Benchmarking of BPML Engines: A Comparison Framework, Statu.	2014	Proceedings of the 20th Inter.	RoelHamerWitz2014	
25	Inproceedings	Hamer et al.	Process Engine Selection Support	2014	On the Slove to Meaningful Int.	Hamer2014	
26	Techreport	Preislinger and Hamer	Static Analysis Rules of the BPML Specification: Tagging, Formalization and Tests	2014		PreislingerHamer2014	
27	Techreport	Roel and Hamer	Testing BPML Engine Performance: A Survey	2014		RoelHamer2014	
28	Inproceedings	Binz et al.	TOSCA: Portable Automated Deployment and Management of Cloud Applications	2014	Advanced Web Services	BinzBrettenboescherKoppEhNZ	
29	Inproceedings	Hamer et al.	Towards a Richness Evaluation Framework for BPML Engines	2014	Proceedings of the 7th IEEE I.	HamerKallisthromKurneEhNZ2014	
30	Inproceedings	Koils and Witz	Towards Application Portability in Platform as a Service	2014	Proceedings of the 8th IEEE I.	KoilsWitz2014	
31	Inproceedings	Preislinger et al.	Towards Standard Conformance BPML Engines: The Case of Static Analysis	2014	Proceedings of the 6th Centr.	PreislingerHamerSchubertEhNZ	

Source jabref.org

## Jabref

- Free and Open Source Software
- Java-based
- BibTeX as native file format
- Export: HTML, Docbook, BibTeXXML, MODS, RTF, Refer/Endnote, and OpenOffice.org as well as LibreOffice
- Import: IS, Medline/Pubmed (xml), Refer/Endnote, INSPEC, BibTeXXML, CSA, ISI Web of Science, SilverPlatter, Scifinder, OVID, Biblioscope, Sixpack, JStor, and RIS
- Firefox add-on for data import from web
- Uses XMP Metadata in PDF