Scientific Working

Chapter I: Scientific Publication and Literature Research

Michael Granitzer

Outline

Scientific Publication and Literature Research

- 1. Scientific Publication Process
- 2. Scientific Citations
- 3. Selecting and Judging Sources
- 4. Research Strategies and Tools

1. Scientific Publication Process

Scientific Publication Process:

Why Publising Results

Definition 1 (Science)

Science is a systematic approach to build and organize knowledge in the form of testable and reproducable explanations and predictions about the world/universe.

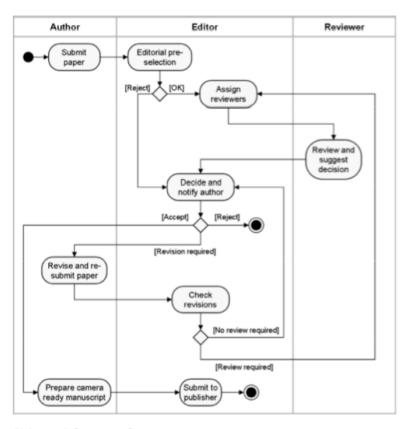
Scientific publications encode the **knowledge gained** through a scientific process. Publications serve multiple purposes¹:

- □ To remember, to understand and to gain new perspectives [BCW03]
- Scientific communication and collaboration [O'c02]
- Retaining intellectual property rights and authorship rights
- Quality control through scientific peer review

Scientific Publication Process:

Publication Workflow and Quality Assurance

Peer Review for quality ensurance



- Siehe auch [Derntl 2002]
- Source: [Der14]

- Author: creator of new knowledge
- Editor: person or group of people finaly deciding to accept/reject a paper
- Reviewer: person creating a recommendation to accept/reject a paper.
- Different kind of reviews: Single Blind, Double Blind, Open

Scientific Publication Process:

Quality Indicators

- □ **Kind of Publication:** Book, Journal, Conference Proceedings, Workshop Proceedings, Technical Reports, Web
- Author, Editor: Seniority, authority on a field, past publications, curiculum vitae
- □ **Publisher:** ACM, IEEE, Springer, Open Acess Sources
- □ **Citations:** Citations by other papers (quantity, quality), uptake in science and industry

Citations are an important source for judging the quality of an indicator. However, citations only provide a rough guideline. Always judge potentially interesting content on your own.

What are citations

Definition 2 (Citations)

Citations reference published or unpublished sources from a marked text passages such that the source is uniquely identifiable from text.

Purpose

- Concise Publications: Refer to additional background material not explained in the publications but relevant
- □ **Contribution:** Contrast your work with related work in the field to show the new contribution.
- Acknowledgment: Acknowledging past work that enabled / influenced the new contribution

How to cite?

Two elements:

The Link: Marked Text Passage for uniquely identifying the context of a citation and linking the text passage to the bibliographic metadata.

and Zaragoza (Metzler and Zaragoza 2009) introduced semi-parametric and non-parametric weighting approaches in addition. In particular, they extend Anh and Moffats (Anh and Moffat, 2005) approach **Bibliographic Metadata** for uniquely identifying the source in a (online) library, usually contained in a so called reference section at the end.

REFERENCES

Anh, V. N. and Moffat, A. (2005). Simplified similarity scoring using term ranks. In SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pages 226–233, New York, NY, USA. ACM.

Bishop, C. M. (1996). Neural networks for pattern recognition. Oxford University Press, Oxford, UK.

Blinded. Blinded. In International Conference on Database and Expert Systems Applications DEXA 'XX. IEEE.

Formatting "the Link"

- □ Numerical: e.g. [1]
- □ Harvard Style^a: (Autors, Year)
 - Use and or & for more authors,
 e.g. (Helic & Strohmaier 2008)
 - Use et al. (latin for "and others") in case of more than 2 authors, e.g. (Granitzer et al. 2008)
 - Use n. d. (abbr. for "No Date") in case there exists no date, e.g. (Smith n.d.)
- □ Alternative Abbrevations for Authors by using first character of surname, e.g. (AM05),(DSW+06)
- □ Footnote can be used too, but quite uncommon in computer science

ahttp://libweb.anglia.ac.uk/
referencing/harvard.htm

Knowledge acquisition still remains a bottle neck for most knowledge based applications as well as in the Semantic Web. However, the large number of unstructured text information in the digital universe [1] suggests itself to be exploited by automatic means in order to extract concept, relations and subsequently ontologies [6]. Focusing on taxonomies, efficient automatic methods for the English

and Zaragoza (Metzler and Zaragoza, 2009) introduced semi-parametric and non-parametric weighting approaches in addition. In particular, they extend Anh and Moffats (Anh and Moffat, 2005) approach

REFERENCES

Anh, V. N. and Moffat, A. (2005). Simplified similarity scoring using term ranks. In SIGIR '05: Proceedings of the 28th annual invernational ACM SIGIR conference on Research and development in information retrieval, pages 226–233. New York, NY, USA, ACM.

Bishop, C. M. (1996). Neural networks for pattern recognition. Oxford University Press, Oxford, UK.

Blinded. Blinded. In International Conference on Database and Expert Systems Applications DEXA 'XX. IEEE.

References

- John F. Gantz. The expanding digital universe. Technical Report, March 2007.
- Frank Manola and Eric Miller. RDF Primer. W3C, February 2004.
- John Davies, Rudi Studer, and Paul Warren. Semantic Web Technologies: Trends and Research in Ontology-based Systems. John Wiley and Sons, July 2006.

Formatting "the Link"

Some more special cases:

- □ Multiple citations can be packed in one bracket, e.g. [1;2;3], (Granitzer 2008; Helic et al. 2013)
- □ Details in references, e.g. [1, p. 5], (Granitzer 2008, Chapter 1)

Exact details depend on the formatting style depends on the required style by the publisher

Formatting Bibliographic Metadata

- □ Bibliographic metadata should uniquely point to the referenced sources
- It should enable finding the original material in a (online) library
- Major items: Author, Title, Year and Publication Venue
- → Additional items depend on the kind of publication.
 - Conferences: Author, Title, Title of the Proceedings, Year. Optional: Editor, Volume,
 Pages, Month, Publisher
 - Journal/Article: Author, Title, Journal, Year. Optional: Volume, Number, Pages, Month
 - Book: Author or Editor, Title, Publisher, Year. Optional: Number, Series, Edition, Month, ISBN
 - Online Material: Author, Title, Date of Publishing (if available), URL, Date of last visit
 - e.g. Cornell Law School: The Legal Information Institute. Supreme Court Collection. online under http://supct.law.cornell.edu/supct/, last visited 2000-03-11.
 - In order to make URL persistent you should use web archiving platforms like www.webcitation.org

The BibTeX Formats on Wikipedia give a good overview on the bibliographic details².

Different Style Formats in Computer Science

□ ACM - Association for Computing Machinery http:
 //www.acm.org/sigs/publications/proceedings-templates
 □ IEEE - Institute of Electrical and Electronics Engineers http://www.ieee.org/web/publications/authors/transjnl/index.html
 □ Springer - Lecture Notes in Computer Science (LNCS), Lecture Notes in Artifical Intelligence (LNAI) http:
 //www.springer.com/computer/lncs?SGWID=0-164-7-72376-0

Formats in other disciplines:

- □ APA American Psychological Association
- □ AMS American Mathematical Society

When to cite?

You have to cite if you copy text literally or by analogously use of ideas, concepts, methods etc. published elsewhere. Both cases have to be differentiated.

Analogously usage:

- □ Explicitly mentioning the work: e.g. "Anh and Moffat (Anh and Moffat, 2005) argued that binned ranking is a general form of dimensionality reduction."
- □ At the end of the sentence: e.g. "Basically, binned ranking can be seen as a general form of dimensionality reduction (Anh and Moffat, 2005)."
- ☐ At the end of the paragraph, if the whole paragraph relates to others work

Literal Copying Text / Quoting

- □ A literal quote has to be marked accordingly!
- □ Usually one uses double quotes: e.g.
 - As stated in (Hearst et. Al. 2008): "What should the interface differ from what is being offered?"
 - "What should the interface differ from what is being offered?" (Hearst et. Al. 2008)

Concepts, ideas and methods which are well known in a particular domain, do not require citation if used as background material. However, if you build, extend, utilize someones work, you need to reference it (c.f. contributions and acknowledgement purpose above)

When to cite?

Definition 3 (Plagiarism)

"Plagiarism is the *wrongful appropriation* and *stealing and publication* of another author's *language, thoughts, ideas, or expressions* and the representation of them as one's own original work." (Wikipedia)

In case of Scientific Publications:

- □ To disguise someone elses work as your own work
- Missing citations for literal quotes
- □ Copy of words, ideas, methods, concepts etc. without quote
- Incorrect use of bibliographic metadata
- Paraphraseing: syntactically rewriting a text from another source while retaining its semantics

There are different reasons for plagiarism: unknowingly, accidently, not knowing the state of the art, willingly, self-plagiarism

Overview

Finding relevant literature is essential for the quality of scientific work

- □ State-of-the-Art Research for gaining an overview about a field
- Related-Work for embedding your work in the field and finding other approaches solving similar problems / answering similar questions
- □ Avoid Plagiarism

There is no absolute quality criterion for a scientific finding. However, citations provide an indicator

- Citations define how many other researchers use the knowledge provided by an publication

Bibliometrics

Bibliometrics is the field concerned with estimating and interpreting quantitative statistics over citations, authors, institutions and libraries.

Different indices to provide robustness:

- □ **Impact Factor** defines how often articles in a particular source get cited on average ³
- □ Science Citation Index⁴
- □ **h-index**, g-index, hc-index etc.

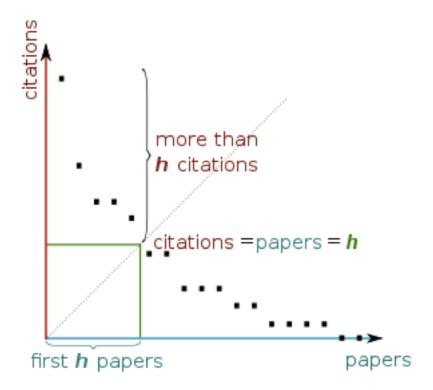
³http://www.sciencegateway.org/impact/

http://scientific.thomson.com/cgi-bin/jrnlst/jloptions.cgi?PC=K

Bibliometrics: Example h-Index

Definition 4 (h-index / Hirsch-Index)

The h-index is the number of h publications that have been cited more than h times



Quality Indicators for a source

Note that errors happen even in scientific articles. You can not read every experiment and redo it, so you need some guidelines for judging the quality of an article (before you read it/use the results)

Judge a source on different quality indicators, not only on statistics (which can be cheated in most cases)

- Bibliometrics
- Acceptance rate on conferences/journals. The lower the better.
- □ Conference/Journal ratings (e.g. CORE Ranking⁵, Conference Ranks ⁶). But beware which organisation provides the ranking and on what data sources.
- Author networks, Expertise of authors (e.g. number of papers published on the topic), relation to expert groups on the field
- □ The number of references in high quality articles
- The peer-reviewing system / length and quality of the reviews
- □ The quality of the reviewer and editors

⁵http://core.edu.au/

⁶http://www.conferenceranks.com/

Good sources in Computer Science

Type of publication: Top conferences > Books, Journals > Conferences > Technical Reports, Web Sources

Top conferences are those with the lowest acceptance rates, a highly renowned programme committee, highest amount of seminal papers

Good sources in Computer science

- □ ACM, especially the ACM special interest groups ACM SIGS (E.g. ACM SIGIR)
- □ IEEE, Elsevier, Springer LNCS/LNAI
- □ The DBLP bibligoraphy (Digital Bibliography & Library Project)
- American Association for Artifical Intelligence AAAI

Beware: A lot of conferences nowadays tend to make businesses out of publications. Usually articles in such venues have lower quality. Tests showed that even automatically generated papers got accepted!

Overview

Searching for research papers is different to web search:

- □ **Exploratory Search:** It is not perfectly clear at the beginning, what you are looking for. You build up knowledge on the domain during search.
- □ Recall matters: Finding the one paper that contains the solution to your problem / answer to your question. In web search a lot of websites provide similar information, so finding the correct one is easier.
- □ **Bibliographic Metadata:** Bibliographic metadata allows you to narrow down the search / enter different search paths
 - Navigate the citation path
 - Search for other papers of the same autor
 - Search for other papers of the same institution
 - Look for other papers in the same conference / journal
 - Look up co-autors and their work
- □ Special Purpose Search Engines

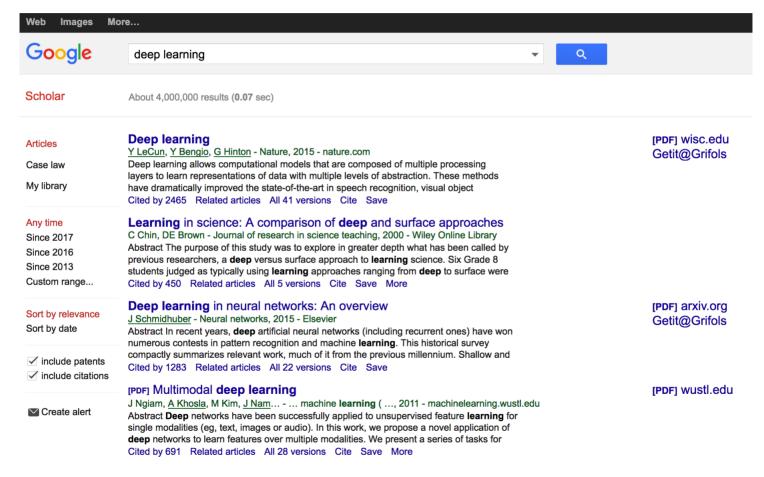
Research Tools

Special Purpose Search Engines

- □ Google Scholar
- □ ACM, IEEE, Springer Portal
- □ Special Purpose Libraries (e.g. the technical library in Hannover)
- □ Elsevier Scopus

Example: Google Scholar Search

Google Scholar provides facets, direct full paper access, citations



Example: Google Scholar Search

Find Similar papers with different keywords via author page



Geoffrey Hinton

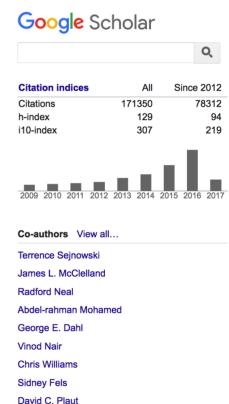
of Toronto &

Emeritus Professor of Computer Science, University of Toronto & Engineering Fellow, Google Inc.

machine learning, neural networks, artificial intelligence, cognitive science, computer science

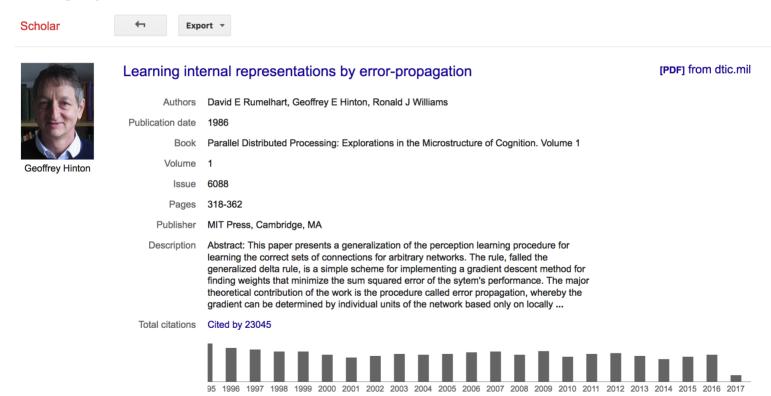
Verified email at cs.toronto.edu - Homepage

Title 1–20	Cited by	Year
Learning internal representations by error-propagation DE Rumelhart, GE Hinton, RJ Williams Parallel Distributed Processing: Explorations in the Microstructure of	23045	1986
Learning internal representations by error propagation DE Rumelhart, GE Hinton, RJ Williams CALIFORNIA UNIV SAN DIEGO LA JOLLA INST FOR	22872	1985
Parallel distributed processing DE Rumelhart, JL McClelland, PDP Research Group IEEE 1, 443-453	21498	1988
Learning representations by back-propagating errors DE Rumelhart, GE Hinton, RJ Williams Nature 323, 533-536	12343	1986
Imagenet classification with deep convolutional neural networks A Krizhevsky, I Sutskever, GE Hinton Advances in neural information processing systems, 1097-1105	11554	2012



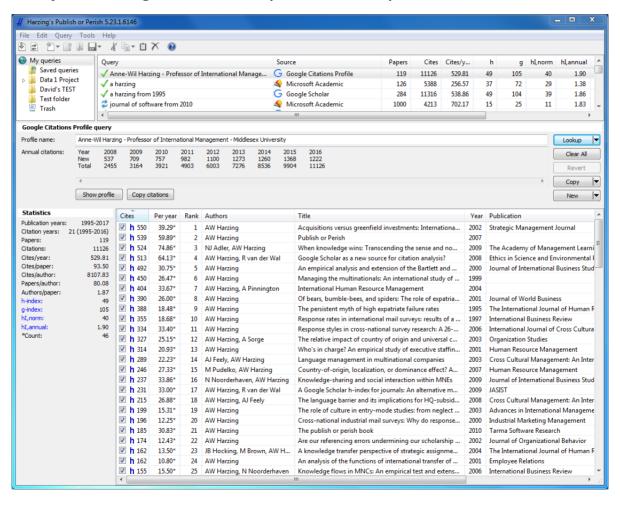
Example: Google Scholar Search

Find papers building on the results or that have been referenced by navigating the citation graph:



Example: Publish or Perish

A tool that analysis Google Scholar queries and provides different indices

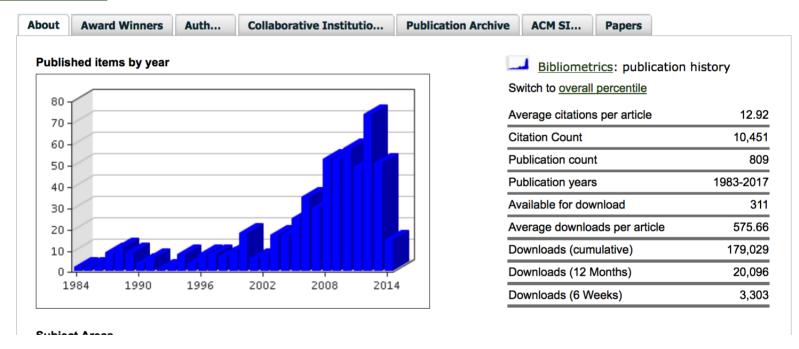


Example: ACM

Institutional profiles

nstitutional Profile: University of Passau

What is an Institutional Profile? BETA



Example: ACM

Institutional profiles

Subject Areas

Document types Formal languages and automata theory Formal Software verification Human computer interaction Human-centered Logic Machine learning Metrics Parallel programming languages Program verification Software development process management Software verification Verification

Most Downloaded Articles (past 6 weeks) / Most Cited Articles

1. Robust and Collective Entity Disambiguation through Semantic Embeddings - 2016

Stefan Zwicklbauer, Christin Seifert, Michael Granitzer

Downloaded 52 times

2. Hypervideos and Interactive Multimedia Presentations - 2017

Britta Meixner

Downloaded 51 times

3. Yesterday, my program worked. Today, it does not. Why? - 1999

Andreas Zeller

Downloaded 42 times

4. Ubiquitous Access to Digital Cultural Heritage - 2017

Christin Seifert, Werner Bailer, Thomas Orgel, Louis Gantner, Roman Kern, Hermann Ziak, Albin Petit, Jörg Schlötterer, Stefan Zwicklbauer, Michael Granitzer

Downloaded 40 times

Tools for Managing your Bibliography

Bibliography: Set of papers relevant for your work|thesis|seminar|...

There are a number of tools that can be used to manage your bibliography

- □ Bibtex/Latex
- □ Word for single articles
- □ Online Tools
 - CiteULike
 - Bibsonomy
 - Google Scholar
 - Zotero
- □ Desktop
 - JabaRef
 - Citavi
- □ Online / Desktop
 - Mendeley

References

- [BCW03] Wayne C Booth, Gregory G Colomb, and Joseph M Williams. *The craft of research*. University of Chicago press, 2003.
- [Der14] Michael Derntl. Basics of research paper writing and publishing. *Int. J. Technol. Enhanc. Learn.*, 6(2):105–123, January 2014.
- [O'c02] Maeve O'connor. Writing successfully in science. Routledge, 2002.
- [Smi90] Alan Jay Smith. The task of the referee. *IEEE computer*, 23(4):65–71, 1990.