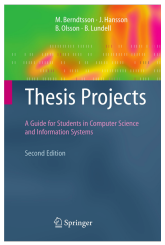
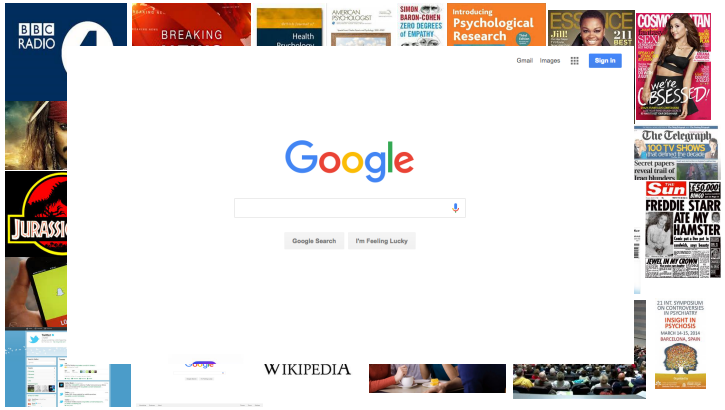
Information gathering: evaluating sources

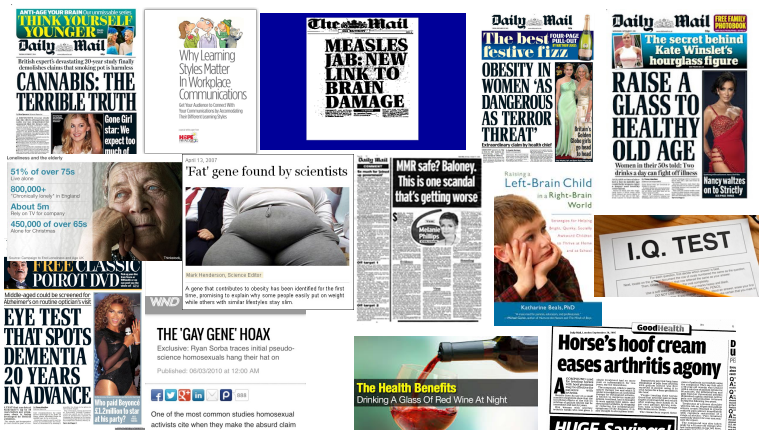
Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. (2007). Thesis projects: a guide for students in computer science and information systems. Springer Science & Business Media.



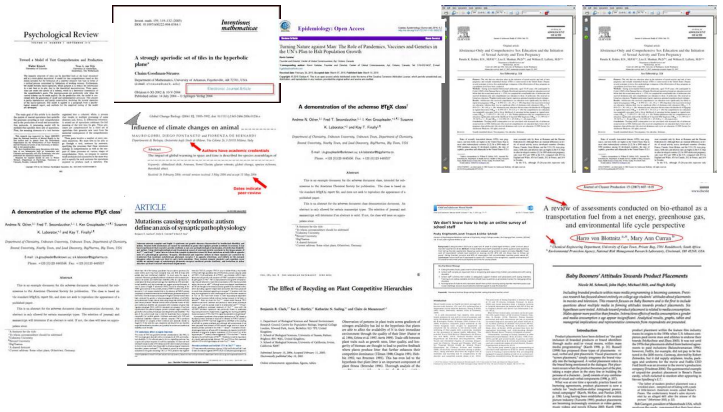
How do we form our knowledge?



Myths?



Facts?



Assessing the source: an "Argument from Expert Opinion"

"Critical Questions" for the Argument from Expert Opinion:

- **Expertise**. How credible is the source?
- **Field**. Is the source expert in the field?
- **Opinion**. What did the source actually say? When?
- **Trustworthiness**. Is the use of the source reliable?
- **Consistency**. Is the source consistent with others? -

(PEER REVIEW!!!)

Scholarly vs Popular

- A scholarly paper is a way for experts in the field to communicate their results/ findings, typically to other experts in the field
- A popular paper (on a research finding) is a communication to the general public about some research

Main academic sources

Conference papers

- Gatherings for researchers to present and discuss their work, typically related to a particular academic discipline and often held at regular intervals
- Output published in "conference proceedings" in the form of conference papers written by the researchers about their work (and often peer-reviewed)
- Types of events: Conference, Workshop, Symposium
  - Check "acceptance ratio"

Main academic sources

Journal articles

- Academic journals are peer-reviewed periodicals in which research relating to a particular academic discipline is published
- Each issue of a journal contains a collection of articles, each article written by a group of researchers
- "Special issues" on a topic

Discipline dependent

- Computer Science is a very dynamic subject, and communication of results, and publication, relies heavily on conferences.
- Other disciplines (e.g. Medical research) publish mainly in journals.

Discipline dependent

- Therefore:
  - A (prestigious) Computer Science conference has a very rigorous peer review process, papers tend to be longer and more exhaustive
  - A Medical conference presents mainly 300ish word "abstracts"
    - the intention is that the full paper will follow on a journal

Publication pipeline

1. Authors submit paper to conference/journal for peer review
   - at least two, could be more reviewers, reading independently
   - for journal, can have many iterations
2. If accepted, the paper is revised by the authors and submitted to conference journal editor
3. The paper is processed to bring it into the publisher's format (typesetting/layout)

Publication pipeline

4. The paper is then included in the publisher's database, made available on-line via the publisher's website, and possibly published in printed form
(not necessarily in that order)

5. Literature databases collect the bibliographic information from several publishers, and add additional information (references with links, citation index) + link back to publisher for full-text of papers

Databases and search engines

The University of Liverpool subscribes to **publisher** DBs
(Access to full-text requires authentication by MWS login and password)

ACM Digital Library	Full-text of all ACM journals and conference proceedings `http://portal.acm.org.ezproxy.liv.ac.uk/dl.cfm`
IEEE Xplore	Full-text of IEEE journals, conference proceedings, and books `http://ieeexplore.ieee.org.ezproxy.liv.ac.uk/`
ScienceDirect	Full-text of Elsevier journals `http://www.sciencedirect.com.ezproxy.liv.ac.uk`
Springer Link	Full-text of Springer journals, conference proceedings, and books `http://www.springerlink.com.ezproxy.liv.ac.uk/`
Wiley Online Library	Full-text of Wiley journals and books `http://onlinelibrary.wiley.com.ezproxy.liv.ac.uk/`

Databases and search engines

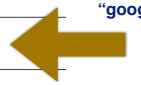
The University of Liverpool subscribes to **literature** DBs
(Access to full-text requires authentication by MWS login and password)

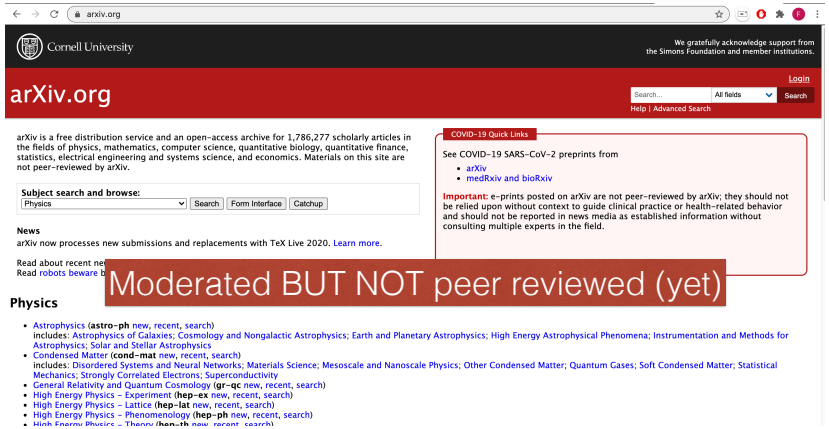
Scopus	Covers 14,000 journals and proceedings series; incl. ACM, Elsevier, IEEE, Springer `http://www.scopus.com/`
Web of Knowledge	Covers 22,000 journals and 192,000 proceedings; incl. ACM, Elsevier, IEEE, Springer `http://apps.webofknowledge.com/`
DISCOVER (UoL)	Meta search engine for ACM Digital Library, IEEE Explore, etc but also Scopus, Web of Science and Google Books `http://www.liv.ac.uk/library/`

Databases and search engines

Freely available (scholarly) sources:

CiteseerX	Digital library of 750k freely available papers in computer and information science `http://citeseer.ist.psu.edu`
Google	General internet search engine `http://www.google.co.uk`
Google Scholar	Searches scholarly literature on the web `http://scholar.google.com`
Scirus	Searches journals (ScienceDirect) and web resources `http://www.scirus.com/`
Microsoft Academic Search	Academic search engine - search academic journals and content for article titles, author names, article abstracts, and conference proceedings `http://academic.research.microsoft.com/`





Lit DBs vs search engines

Lit DBs cover a vast number of academic sources, but

- they do not cover all journals and conferences
- they do not cover books
- they do not cover workshops and similar scientific meetings
- they do not cover technical reports and pre-prints

Web search engines provide much better coverage of all types of publications, but

- typically also return a lot of irrelevant material to a query
- leave it to the user to distinguish high quality from low quality material

Cr*p

... or Craap

The Craap Test

- Developed by Sarah Blakeslee, of the University of California at Chico's Meriam Library
- Useful, multi-platform and multidisciplinary checklist to evaluate sources

- Currency
- Relevance
- Authority
- Accuracy
- Purpose

UNIVERSITY OF LIVERPOOL

COMP 107

Currency

- Currency refers the timeliness of the information
  - When was the information published or posted?
  - Has the information been revised or updated?
  - Is the information current or out-of date for your topic?
  - Are the links functional?

UNIVERSITY OF LIVERPOOL

COMP 107

Relevance

Relevance refers to the importance *for your needs:*

- Does the information relate to your topic/answer your question?
- Who is the intended audience?
- Is the information at an appropriate level (i.e. not too elementary or advanced for your needs)?
- Have you looked at a variety of sources before determining this is one you will use?
- Would you be comfortable using this source for a research paper?

UNIVERSITY OF LIVERPOOL

COMP 107

Authority

Authority refers to the source of the information

- Who is the author/publisher/source/sponsor?
- Are the author's credentials/organisational affiliations given?
- What are the author's credentials/organisational affiliations?
- What are the author's qualifications to write on the topic?
- Is there a contact information, e.g a publisher or e-mail address?
- Does the URL reveal anything about the author or source?

  examples: **.com** (commercial), **.edu** or **.ac.uk** (educational), **.gov** (U.S. government) **.org** (nonprofit organisation), or **.net** (network)

UNIVERSITY OF LIVERPOOL

COMP 107

Accuracy

- Accuracy refers to the reliability, truthfulness, and correctness of the content
  - Where does the information come from?
  - Is the information supported by evidence?
  - Has the information been reviewed or refereed?
  - Can you verify any of the information in another source or from personal knowledge?
  - Does the language or tone seem biased and free of emotion?
  - Are there spelling, grammar, or other typographical errors?

UNIVERSITY OF LIVERPOOL

COMP 107

Purpose

- Purpose refers to the reason the information exists
  - What is the purpose of the information? to inform? teach? sell? entertain? persuade?
  - Do the authors/sponsors make their intentions or purpose clear?
  - Is the information fact? opinion? propaganda?
  - Does the point of view appear objective and impartial?
  - Are there political, ideological, cultural, religious, institutional, or personal biases?

UNIVERSITY OF LIVERPOOL

COMP 107

The CRAAP Test Worksheet

Use the following list to help you evaluate sources. Answer the questions as appropriate, and then rank each of the 5 parts from 1 to 10 (1 = unreliable, 10 = excellent). Add up the scores to give you an idea of whether you should you use the resource (and whether your professor would want you to!).

Currency: *the timeliness of the information*
- When was the information published or posted?
- Has the information been revised or updated?
- Is the information current or out-of date for your topic?
- Are the links functional?

Relevance: *the importance of the information for your needs*
- Does the information relate to your topic or answer your question?
- Who is the intended audience?
- Is the information at an appropriate level?
- Have you looked at a variety of sources before choosing this one?
- Would you be comfortable using this source for a research paper?

Authority: the source of the information
- Who is the author/publisher/source/sponsor?
- Are the author's credentials or organizational affiliations given?
- What are the author's credentials or organizational affiliations given?
- What are the author's qualifications to write on the topic?
- Is there contact information, such as a publisher or e-mail address?
- Does the URL reveal anything about the author or source?

Accuracy: the reliability, truthfulness, and correctness of the content
- Where does the information come from?
- Is the information supported by evidence?
- Has the information been reviewed or refereed?
- Can you verify any of the information in another source?
- Does the language or tone seem biased and free of emotion?
- Are there spelling, grammar, or other typographical errors?

Purpose: the reason the information exists
- What is the purpose of the information?
- Do the authors/sponsors make their intentions or purpose clear?
- Is the information fact? opinion? propaganda?
- Does the point of view appear objective and impartial?
- Are there political, ideological, cultural, religious, institutional, or personal biases?

45 - 50 Excellent | 40 - 44 Good
35 - 39 Average | 30 - 34 Borderline Acceptable
Below 30 - Unacceptable

Total:

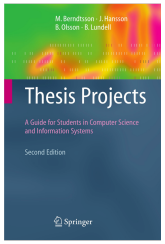
- Give a score 1/10 to each category
- Above 40, good or excellent
- Below 30, unacceptable

UNIVERSITY OF LIVERPOOL

COMP 107

HOW TO REFERENCE

Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. (2007). Thesis projects: a guide for students in computer science and information systems. Springer Science & Business Media.
Chapter 6



UNIVERSITY OF LIVERPOOL

COMP 107

References

A **reference** is a **description** that identifies an information **source**.

Purpose of References

- Show how work extends the current state-of-the-art knowledge in the area
- Prove originality of work
- Give credit to other people's work (avoid accusations of plagiarism)
- Support and validate arguments made
- Demonstrate familiarity with work done in the area

The culture of citing

- Fiction writing: There might be sources for your work, but these are never acknowledged, unless they serve a literary purpose.
- Journalism: There must be sources, but these are rarely acknowledged (sometimes for a good reason, often for no reason)
  - Exceptions:
    – Quotation of something said in public
    – Reports (government, research, 'think tank')
  - But acknowledgements are not always in a format considered acceptable in academia
  - E.g.: https://www.theguardian.com/science

The culture of citing

- How many entries does the bibliography of the following textbooks contain?
  - W. Hughes, J. Lavery, and K. Doran: Critical Thinking: An Introduction to the Basic Skills (6th revised edition). Broadview Press, 2010. **none!**
  - R. Morelli and R. Walde: Java, Java, Java: Object-Oriented Problem Solving (3rd edition). Pearson, 2006. **none!**
  - R. Elmasri and S. Navathe: Fundamentals of Database Systems (6th Edition). Addison-Wesley, 2010. **LOADS!**

Discipline dependent

- In Academic writing all sources that make a contribution to your work must be acknowledged
- Styles can be different from discipline to discipline, e.g.
  - ***Philosophy***:

    A lot of quotations, meticulously indicated as such. A lot of discussion/argumentation with reference to previous work ; extensive and precise referencing

Discipline dependent

- **Mathematics**:
  Important concepts and results are given specific names, (often the name of the first person to introduce the concept/result, often these concepts/results are considered common knowledge)
  - A lot of concepts have a fixed definition, results have a specific wording, so definitions and results are stated or used without indication of a source;
  - verbatim copying and close paraphrasing of definitions and results is standard (but not that of proofs!)

Rules of Thumb

- If you use words or ideas from any document/medium, even produced by yourself, then the source must be cited
- If you gain words or ideas through conversation, written or spoken, then the source must be cited
- If you use the exact words/phrase from any document, medium or conversation, it must not only be cited but also indicated as **quotation** (“”)
- If you reproduce audio-visual materials (with permission), then the source must be cited

No citation required if:

- you are writing about your own experiences, your own thoughts, your own observations and insights, and your own conclusions **that have not been published before**
- you are writing about your own work and your own experiments **that have not been published before**
- you are reusing your own audio-visual materials
- you are using common knowledge or generally accepted facts
- in the context of student submissions: you are using facts (but not exact words) from recommended textbooks

How to refer to a source

- General principle: just by looking at the reference, without retrieving the actual source, a knowledgeable reader must collect all information about the source
  - the **structure** of the reference, and the information included, will give an indication of the **type** of source

Type of sources

- Author's own past work
- Private communication with others
- Newspaper articles, Radio or TV programmes
- Web pages
- Books or book chapters
- Research publications (conference papers, articles in journals)
- Theses (Masters, MPhil, PhD)
- Formal documents: Publications, Legislation/Court cases, Patents, Manuals…

URLs are not enough

- Because the URL might not be valid in the future
- Because a reader knowledgeable in the subject area should be able to identify a source without the need to retrieve it
- Because the reputation of a journal / conference is often taken as a proxy indicator of quality
- URLs should only be given in addition to the required bibliographic information, never **instead** of it

Digital Object Identifiers

- The Digital Object Identifier (DOI) System is trying to solve the problem of impermanence of URLs
  - a unique and permanent identifier, DOI name is assigned to a source, the location is stored separately, and only accessible via a **DOI resolver**, e.g. http://dx.doi.org
- DOIs are preferable over alternative URLs, but:
  - Still, a reader knowledgeable in the subject area should be able to identify a source without the need to retrieve it
  - Still, the reputation of a journal / conference is often taken as a proxy

Type/content of the reference

- What information is required about where the work can be obtained depends on its type
- The information provided on where the work can be obtained indicates its type
  - Book:
    Herman T. Tavani. 2010. Ethics and Technology: Controversies, Questions, and Strategies for Ethical Computing (3rd ed.). Wiley Publishing.
  - Journal Paper:
    Herman T. Tavani. 2011. Can we Develop Artificial Agents Capable of Making Good Moral Decisions?. Minds Mach. 21, 3 (August 2011), 465-474. DOI=http://dx.doi.org/10.1007/s11023-011-9249-8

How to refer to a…

Book

- Author(s) or editor(s)
- Title and subtitle
- Edition, if not the first, for example 2nd ed.
- Series and individual volume number (if any)
- Publisher
- (Place of publication)
- Year of publication

UNIVERSITY OF LIVERPOOL
COMP 107

How to refer to a…

Chapter/section of a book with separate authors

- Author(s) of the chapter/section
- Title and subtitle of the chapter/section
- Author/editor of collected work
- Title and subtitle of collected work
- Chapter/section referred to
- Page numbers of chapter/section referred to
- Publisher
- (Place of publication)
- Year of publication

UNIVERSITY OF LIVERPOOL
COMP 107

How to refer to a…

Conference proceedings

- Editor(s) of proceedings (if information is available)
- Name and number of conference
- Location of conference (if appropriate)
- Time of conference
- Title of published work; if different from the name of the conference
- Series and individual volume number (if any)
- Publisher
- Place of publication
- Year of publication

UNIVERSITY OF LIVERPOOL
COMP 107

How to refer to a…

Conference paper

- Author(s) of the paper
- Title and subtitle of the paper
- All information on the conference proceedings plus
- Page numbers of the paper

How to refer to a...

Journal article

- Author(s) of the article
- Title and subtitle of the article
- Title of the journal
- Volume and part number
- Page numbers of article
- Date, month or season of the year, if appropriate
- Year of publication

Note: Information on the publisher is typically not required

UNIVERSITY OF LIVERPOOL
COMP 107

How to refer to a...

Newspaper / Magazine article

- Author(s) of the article
- Title and subtitle of the article
- Name of the newspaper / magazine
- Volume and part number
- Page numbers of article
- Date, month or season of the year, if appropriate
- Year of publication

- Information on the publisher is typically not required
- If no author is indicated, leave this information out
  when referring to the article use the name of the newspaper as author

UNIVERSITY OF LIVERPOOL
COMP 107

How to refer to a...

1. One-off or regularly recurring TV programmes
   - Report title, if applicable
   - Programme Title
   - Name of channel
   - Date of transmission
   - Time of transmission
2. Episode of a TV series
   - Writer(s), if relevant
   - Director(s), if relevant
   - Episode title
   - Producer(s), if relevant
   - Series title
   - Season and episode
   - Name of channel
   - Date of transmission
   - Time of transmission

UNIVERSITY OF LIVERPOOL
COMP 107

How to refer to a...

Thesis and dissertation

- Author of the work
- Title and subtitle of the work
- Type of work
- Awarding institution including its address
- Year, possibly month, of publication

How to refer to a…

Web pages

- Author(s) of the web page(s)
- Title and subtitle
- URL
- Date (possibly time) of last modification, if available
- Date (possibly time) of access

How to refer to a…

Legislation

- Title of the act
- Year in which the act came into force
- URL where the act can be found
- Date on which you have accessed that URL

Court cases

- Case name, typically given in the form claimant v defendant
- Year in which the case concluded
- Court
- Case/judgement number or document details
- URL where the document can be found
- Date on which you have accessed that URL

👍 OR 👎

Bad:
JAVA, JAVA, JAVA by Ralph Morelli

Good:
Ralph Morelli: Java, Java, Java: Object-Oriented Problem Solving, 2nd edition. Prentice Hall, 2003.

👍 OR 👎

Bad:
Marco Dorigo and Thomas Stutzle, Ant Colony Optimization.

Good:
Marco Dorigo and Thomas Stützle: Ant Colony Optimization. Bradford Book, 2004.

👍 OR 👎

Bad:
Marco Dorigo, Gianni Di Caro, Michael Samples, Ant Algorithms, third international workshop, Ant 2002, Brussels, Belgium, September 2002, Proceedings.

Good:
Marco Dorigo, Gianni Di Caro, and Michael Samples, editors: Ant Algorithms: Third International Workshop, ANTS 2002, Brussels, Belgium, September 12–14, 2002, Proceedings. Lecture Notes in Computer Science 2463. Springer, 2002.

👍 OR 👎

Bad:
`http://www.cut-the-knot.org/blue/Stern.shtml`

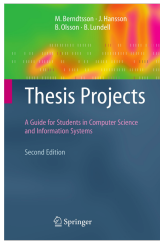
Good:
Alexander Bogomolny: Stern-Brocot Tree.
`http://www.cut-the-knot.org/blue/Stern.shtml`.
Last modification 17 June 2000. Accessed 26 October 2013.

Come on get real: I'll never remember this!

- Google (Scholar) is your friend!
- https://scholar.google.co.uk/

HOW TO CITE

Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. (2007). Thesis projects: a guide for students in computer science and information systems. Springer Science & Business Media.
Chapter 6



Reference vs Citation

- A reference is a description that identifies an information source
- A citation is the use of a reference in the text.
  - Not a good practice to simply list all your references at the end of your document
  - Need to clearly mark what is your own work and what is the work of others
  - Need to help readers get more info on what they are interested in

Where to place a citation

- Before a punctuation mark:
  The human brain contains approximately 50 billion neurons (Smith, 1994).
- At a logical place in a sentence:
  According to some researchers (Smith, 1994) there are 50 billion neurons in the human brain.
- At a grammatically correct place in a sentence:
  According to Smith (1994), there are 50 billion neurons in the human brain.

Where to place a citation

- Before a list of items
  There are five categories of users (Anderson, 2008): (1) students, (2) teachers, (3) professors, (4) technical staff, (5) administrative staff.
- Following quotations
  "In the experiments it is shown that the human brain has 50 billion neurons. Many of the types of neurons have yet to be classified. We strongly encourage other researchers to develop tools and techniques that will assist the process of categorising the neurons." (Smith, 1994, p. 345)

👍 OR 👎

- *Recent work has reported that the importance of computers in industry cannot be overestimated. Several useful services (such as booking train tickets) rely on computers. However, the importance of using computers in our everyday life has been questioned. It has been argued that having too many computers in our everyday life causes security problems, since people cannot protect their computers from hackers and Internet viruses. The researchers are still debating these hot topics. (Jones, 1993)*
- The placement suggests that the whole paragraph is taken from Jones 1993 (as a paraphrase), in particular the last sentence.
  - Better: *Recent work (Jones, 1993) has reported that…*

👍 OR 👎

- *For a long time, the best stock market predictions have been achieved by the Epsilon neural network architecture (Myers and Sang, 1997, Niven, 1999).*
- Who developed the architecture? Who applied it to stock market predictions?
- Better: *The Epsilon neural network architecture, proposed by Myers and Sang (1997), has for a long time been the most accurate method for stock market prediction (Niven, 1999).*

Bibliography style

- A bibliography style determines
  - how citations are presented
    - what information is in a citation
    - what a citation looks like where the citation is placed
  - how references are presented
    - order of references within a bibliography
    - order of information within a reference fonts
    - punctuation

3 main classes of bib style

1. Ordinal Number

- Sources listed in the bibliography are sorted according to some ordering, typically based on the authors' names, and numbered consecutively
- Citations in the text are given as (lists of) numbers cross-referencing the bibliography, enclosed in square brackets

Ordinal number bib style

Example:

Key techniques for utilising temporal logic specifications have been investigated, including verification via proof [3] and verification via model-checking [1,2].

Bibliography

1. E. Clarke, O. Grumberg, and D. A. Peled: Model Checking. MIT Press, 2000.
2. K. L. McMillan: Symbolic Model Checking. Kluwer, 1993.
3. M. Vardi and P. Wolper: Reasoning about infinite computations. Information and Computation 115:1–37, 1994.

3 main classes of bib style

2. Author-Date

- Sources in the reference list are arranged alphabetically by the authors' surnames (phone book order);
- work by the same authors are arranged by year of publication, starting with the earliest;
- more than one work with the same authors and date, a letter is added to the year of publication to distinguish them
- The year of publication typically immediately follows the list of authors

Author-Date bib style

- Example:

Bibliography

E. Clarke, O. Grumberg, and D. A. Peled (2000). Model Checking. MIT Press.

K. L. McMillan (1993). Symbolic Model Checking. Kluwer.

P. Wolper (1996a). Where is the Algorithmic Support? ACM Computing Surveys 28(4):58.

P. Wolper (1996b). The Meaning of "Formal". ACM Computing Surveys 28(4):127.

Author-Date bib style

- A citation is given by the authors' names and the date enclosed in parentheses unless the authors' names are part of the sentence, for example:
  - *While Wolper (1996a) states that he does not argue that compositionality in proof systems for concurrency is undesirable, he claims that achieving it without algorithmic support is mostly useless.*
  - *Recent work (Wolper, 1996a, 1996b) stresses the importance of algorithmic support for formal methods.*
  - *Wolper (1996a, 1996b) stresses the importance of algorithmic support for formal methods.*

3 main classes of bib style

3. Abbreviation

- Mix of ordinal number style and author-date style
- Sources in the bibliography are presented like in ordinal number style, but instead of numbering them, each source is given a unique identifier based on authors' names and year of publication, with additional letters to disambiguate duplicate abbreviations

Abbreviation bib style

- Example:

Bibliography

[CGP00] E. Clarke, O. Grumberg, and D. A. Peled. Model Checking. MIT Press, 2000.

[vdG94] R. A. van der Goot. Strategies for modal resolution. Master's thesis, Delft University of Technology, The Netherlands, 1994.

[Wol96a] P. Wolper. Where is the Algorithmic Support? ACM Computing Survey 28(4):58, 1996.

[Wol96b] P. Wolper. The Meaning of "Formal". ACM Computint Survey 28(4):127, 1996.

Abbreviation bib style

- Citations are given by using the references, in brackets:
  - *Key techniques for utilising temporal logic specifications have been investigated, including verification via proof [VW94] and verification via model-checking [CGP00,McM93].*
  - *Recent work [Wol96a, Wol96b] stresses the importance of algorithmic support for formal methods.*
  - *Wolper in [Wol96a,Wol96b] stresses the importance of algorithmic support for formal methods.*

👍 OR 👎

- A citation in ordinal-number style never starts a sentence
  👎: *[9] Disaster rescue is a serious social issue.*
  👍: *Disaster rescue is a serious social issue [9].*
- In Computer Science publications a citation never comes after the end of sentence except for quotations
  👎: *2-on-2 teams of autonomous mobile robots play games in a rectangular field color-coded in shades of grey. [9]*
  👍: *2-on-2 teams of autonomous mobile robots play games in a rectangular field colour-coded in shades of grey [9].*

👍 OR 👎

- A citation never occurs in a section heading:
  👎: *Section 5: The History of Robocup (Henry 2006).*
- In ordinal-number style a list of citations is a comma-separated list of ordered numbers enclosed in one pair of square brackets
  👎: *The humanoid soccer robots are fully autonomous [5] [9].*
  👎: *The humanoid soccer robots are fully autonomous [9,5].*
  👍: *The humanoid soccer robots are fully autonomous [5,9].*