Lecture 2 - Information search and evaluation

Christoph Kessler
Agenda

- Information / literature search
- Publication types and quality assessment
- Plagiarism and copyright
- Proper referencing
- Reference management
- Background literature vs. Related work
## Engineering information vs Scientific information

<table>
<thead>
<tr>
<th></th>
<th>Engineering</th>
<th>Science</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Questions</strong></td>
<td>How to solve a problem</td>
<td>How to explain something</td>
</tr>
<tr>
<td><strong>Reliability</strong></td>
<td>Working solutions, proven theories</td>
<td>Cited work</td>
</tr>
<tr>
<td><strong>Sources</strong></td>
<td>White papers, software projects, reviewed publications</td>
<td>Reviewed publications</td>
</tr>
</tbody>
</table>
Finding information / literature

Learn about the subject area:
• use Wikipedia, books, survey articles and previous course material
• Extract *keywords* that you can use when searching papers.
• Use Google Scholar and Unisearch first, specific publications second

Search articles and books

[Google Scholar]

[Unisearch]
Wikipedia??

- Not stable
- Anonymous authors
- May be biased
- May be incomplete
- Varying quality
- Unclear quality control

→ NOT CITEABLE!

But still useful for getting an early overview of a new topic area
- by following (and reading) given references
- finding appropriate keywords and their synonyms for better search

Software design pattern

From Wikipedia, the free encyclopedia

In software engineering, a design pattern is a general reusable solution to a commonly occurring problem within a given context in software design. A design pattern is not a design that can be transformed directly into source or machine code. It is a description of a template for how to solve a problem that can be used in many different situations. Patterns formalized best practices that the programmer can use to solve common problems when designing an application or system. Object-oriented design patterns typically show relationships and interactions between classes or objects, without specifying the final application of these objects that are involved. Patterns that imply mutable state may be unsuited for functional programming languages, some patterns can be rendered unnecessary in languages with built-in support for solving the problem they are trying to solve, and object-oriented patterns may not necessarily suitable for non-object-oriented languages.

Design patterns may be viewed as a structured approach to computer programming between the levels of a programming paradigm and a concrete algorithm.

Contents [hide]
1 Types
Stå på giganters axlar
How to find information

"HLA active probing runtime performance requirements in a Wide Area Network"

There is no related work!?!"
Gathering additional keywords and synonyms

High Level Architecture

From Wikipedia, the free encyclopedia

This article is about the IEEE standard for distributed simulation. For other uses, see High- and low-level.

The High-level architecture (HLA) is a standard for distributed simulation, used when building a simulation for a larger purpose than the simulation itself. The standard was developed in the 90's under the leadership of the US Department of Defense and was later translated into an international IEEE standard. It is a recommended standard within NATO through STANAG 4603. Today the HLA is used in a number of defense and security and civilian applications.

The purpose of HLA is to enable interoperability and reuse. Key properties of HLA are:

- The ability to connect simulations running on different computers, locally or widely distributed, independent of their operating systems, into one Federation.
- Ability to specify and use information exchange data models, Federation Object Models (FOMs), for different application domains.
- Services for exchanging information using a publish-subscribe mechanism, based on the FOM, and with additional filtering options.
- Services for coordinating logical (simulation) time and time-stamped data exchange.
- Management services for inspecting and adjusting the state of a Federation.
Smart Searching

HLA simulation
active probing fault detection and localization
runtime performance requirements latency, throughput
in a Wide Area Network IP networks
Iterative search

Recursive search through the world-wide digraph of publications in both directions – following the “cited-by” relation
“This paper by author X was right on target. Did X maybe write other papers about this or related topics?”

- Author homepage (if existing)
- Google Scholar
- DBLP: https://dblp.uni-trier.de
Scientific publishing

- Primary studies
- Secondary studies
- Text books

“What”
Scientific publishing

Peer-reviewed publications
- Conference proceedings
- Journal papers

Non-reviewed publications
- Technical reports
- White papers

"How"
Publication types

can often be inferred already from a descriptive title alone – without even accessing the paper!


*Position paper or survey paper*, no new research results. Shares experience on software architecture research and development. Published in a journal.
Publication types


*Systematic Literature Review*, secondary study. Published at a conference.
Publication types


*Guidelines textbook on empirical methods in Software Engineering.*
Publication types


**Technical report**, non-reviewed publication.

No empirical support for claims, but suggestions of an architecture.

Older technical reports can be hard to get hold of.

Find a published (conference, journal, book chapter) version of this work that can be cited instead.

**Case study,**

*reviewed publication in journal*
Publication types


_**Experimental study,**_
peer-reviewed publication presented at a conference and published in proceedings from the conference.
What are results?

<table>
<thead>
<tr>
<th>Type</th>
<th>How?</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedure/technique</td>
<td>Formal proofs, experiments, statistical support</td>
<td>Selection of benchmark programs, Proper use of statistics</td>
</tr>
<tr>
<td>Descriptive Models</td>
<td>Experience reports: Interviews, observations, usage data</td>
<td>Properly accounting for reality</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Real systems and people</td>
</tr>
</tbody>
</table>

Many (successful) papers contain an explicit list of **contributions** at the end of the introduction section, which include the main results.
What are strong results?

Real systems and proper analysis

<table>
<thead>
<tr>
<th>Type of validation</th>
<th>Submitted</th>
<th>Accepted</th>
<th>Ratio Acc/Sub</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>48 (16%)</td>
<td>11 (26%)</td>
<td>23%</td>
</tr>
<tr>
<td>Evaluation</td>
<td>21 (7%)</td>
<td>1 (2%)</td>
<td>5%</td>
</tr>
<tr>
<td>Experience</td>
<td>34 (11%)</td>
<td>8 (19%)</td>
<td>24%</td>
</tr>
<tr>
<td>Example</td>
<td>82 (27%)</td>
<td>16 (37%)</td>
<td>20%</td>
</tr>
<tr>
<td>Some example, can't tell whether it's toy or actual use</td>
<td>6 (2%)</td>
<td>1 (2%)</td>
<td>17%</td>
</tr>
<tr>
<td>Persuasion</td>
<td>25 (8%)</td>
<td>0 (0.0%)</td>
<td>0%</td>
</tr>
<tr>
<td>No mention of validation in abstract</td>
<td>84 (28%)</td>
<td>6 (14%)</td>
<td>7%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>300 (100.0%)</td>
<td>43 (100.0%)</td>
<td>14%</td>
</tr>
</tbody>
</table>

How to evaluate papers

- Relevance = \( f(\text{title, year, abstract, venue, } \#\text{citations}) \)
  - The more specific the paper, the fewer citations?
- Literature reviews: meta studies
- Publication types: journals, conferences, book chapters
- Trust is inherited from the journal’s, conference’s or publisher’s reputation
  - Web-of-Science/Scopus listed journals, Norwegian level ranking, ...
    see liu.se/en/libraryguides-and-support
  - CORE ranking of conferences and journals www.core.edu.au
  - Beware of predatory publishers, fake conferences and fake journals
    – check Beall’s list if unsure beallslist.weebly.com
- Read with critical eyes.
  Refer to the real main results of the paper (describe in your own words), not what the authors have themselves written in the abstract or introduction
What about white papers/other stuff?

White-papers, data sheets, manuals, web pages, blog entries etc.?

• Not peer-reviewed
• URL may change within a few years, or the paper be removed/replaced.
• Use to support existence: ”There are several implementations of Flux controllers”
• Not to support claims and propositions: ”Flux controllers are more user friendly than Flax controllers”
• Web pages, blog entries and similar volatile material: Specify timestamp of last access.
  – Some examiners prefer to have web references as footnotes rather than cited references

ArXiv / CoRR?

• These are usually preprints – may not be peer-reviewed in this form (exceptions exist)
  – stable URL, but everyone can upload a paper to ArXiv – thus less trustworthy than published material
  – If published, cite the journal / conference version instead.
Plagiarism and copyright
NoPlagiat: self-study tutorial for avoiding Plagiarism and Copyright Issues

Self-study Tutorial

As part of the library's work with plagiarism issues, we have set up a self-study tutorial designed to give you a better idea of what plagiarism is and how to avoid it. In the first hand the tutorial is designed for undergraduate students, but all are free to make use of it.

There are five modules to work through (choose an option via the menu). If you are beginning your studies at LiU or unsure about what plagiarism is then we recommend that you begin with "What is plagiarism?" and work your way through the modules in order. The tutorial takes about 30 minutes to complete. If you are after specific advice on a particular element then it also works to jump straight to that.

Within each module there is some background information, some examples and then a few exercises to test whether you understood or not.

At any point, if you have a question, feel free to contact us: plagiering@bibl.liu.se

Start the tutorial
Using image without reference

Plagiarism + copyright violation

Using image with reference

Copyright violation

Using own/CC image with reference

OK!
Damages claimed for pictures in degree project

A LiU alumnus risks having to pay SEK 4,000 for copyright infringement concerning pictures used in a degree project he did eleven years ago.

The work dealt with how to identify various deciduous trees in winter, and the text was illustrated with 12 photos and 12 drawings from an existing book on the subject. The project was approved in 2000.

“The teaching student provided the book’s title under the illustrations, but not the author’s name nor had he asked the author for permission to use the pictures,” says Martin Putsén, a lawyer with the University administration.

It could now cost the former student SEK 4,000. The author has claimed copyright infringement damages through SLFF, the Swedish Association of Educational Writers, after discovering the unlicensed use. SLFF has also demanded that the degree project no longer be made available in the University library.
Using references properly
Citing the right reference

"Software product lines are related software products that are customized to different customers [1]."


Not the main result of [1]


Cite the book by Pohl et al. instead [2]
References

There are a number of conventions of how to use references properly: use in-text references or outside-text references consistently. IEEE has a good standard for this.

Odersky et al. have studied software design patterns [1].

Odersky et al. (2010) have studied software design patterns.
Over a quarter of the ICSE 2002 abstracts give no indication of how the paper's results are validated, if at all [1].

4.2 Which of these are most common?

Alas, well over a quarter of the ICSE 2002 abstracts give no indication of how the paper's results are validated, if at all. Even when the abstract mentions that the result
Quotations and Citations

Bansiya and Davis claim that the QMOOD model may address “different weightings, other perspectives, and new goals and objectives” [1].

3.8 Refining and Adapting the Model
The QMOOD quality model allows changes to be easily made to the model to address different weightings, other perspectives, and new goals and objectives. At the lowest

Citations, stylistic detail: *et al.*

- One author: ... see Miller [23].
- Two authors: ... see Anderson and Brown [24].
- Three or more authors: ... see Young *et al.* [25].

*et al.* is a Latin abbreviation for “et alii” (and others).
- similarly:  *e.g.* = *exempli gratia* (for the sake of an example)
  *i.e.* = *id est* (that is)
  *etc.* = *et cetera* (and the remaining ones)
Managing references
Fast Crown Scheduling Heuristics for Energy-Efficient Mapping and Scaling of Moldable Streaming Tasks on Manycore Systems

Manycore CPU core allocation to moldable parallel actor tasks is restricted to subsets defined by hierarchical decomposition of the set of cores --> sizes: powers of 2. Combines core allocation, mapping and DVFS for tasks. Fast heuristics and ILP solution.
BibDesk

- BibTeX frontend
- Open-source
- MacOS
Mendeley

System for reference management and sharing
Some Advice
Writing about what you have read

• Keep a copy (hard or soft) of all papers etc. that you have read
  – Create a BibTeX entry when finished

• Take notes of what you have read
  – Maybe in the `annotate={...}` field of the paper’s entry in your BibTeX file(s)
  – Summarize the main insights with your own words

• Consider what needs to be in your thesis.
  Do not write everything you have read into your report.
  Remember to keep a strong connection to your main method/results.
Background Literature vs. Related Work?

- **Related work**: Studies that are similar in style and objective to what you are doing in your own thesis work, i.e., work that you can compare to.
  - Published work addressing the same or a slightly different problem, e.g., for a different processor architecture, programming language, or for a variant of the algorithmic problem considered.
  - Read, digest, describe, compare. At least 1 paragraph per paper.
  - Relate your own experimental evaluations and conclusions to those works and explain the similarities and differences, strengths and weaknesses, thereby leading to a proper related work discussion chapter in your thesis and possibly stronger conclusions.

- Manuals, survey articles, books etc. for theory, systems and techniques that you use for your work are background literature, not related work (but should nevertheless be cited).
Discussing and Comparing Related Work

- **Classify** related approaches and **compare** their properties, assumptions and results with each other and with yours.

- Suggestion: Make a **feature synopsis table**
  - Shows immediately where / how your solution outperforms the competitors.
  - Helps classifying and grouping the related work description into sections.
  - Great to have in an overview or summary section in the Related Work chapter.
  - Great to have on a (backup) slide in the final presentation.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Levin et al. [17]</th>
<th>Sun et al. [38]</th>
<th>Köhler et al. [13]</th>
<th>Ours (real)</th>
<th>Ours (synthetic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthetic/Real</td>
<td>Synthetic</td>
<td>Synthetic</td>
<td>Real</td>
<td>Real</td>
<td>Synthetic</td>
</tr>
<tr>
<td>Blur Model</td>
<td>Uniform</td>
<td>Uniform</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Both</td>
</tr>
<tr>
<td>Latent Images</td>
<td>4</td>
<td>80</td>
<td>4</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>Kernels / Trajectories</td>
<td>8</td>
<td>8</td>
<td>12</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>Blurred Images</td>
<td>32</td>
<td>640</td>
<td>48</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>Depth variation</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Evaluation</td>
<td>PSNR/SSIM</td>
<td>PSNR/SSIM</td>
<td>PSNR</td>
<td>User study</td>
<td>User study</td>
</tr>
</tbody>
</table>

Summary

• Start learning about the subject, then find proper support for your claims. Use different sources for learning and as references to support specific claims.
• Plan sufficient time for literature search and review, especially but not only in the first weeks of your project. Keep a “reading list” of found papers, read regularly during the project, and take notes immediately.
• There are different types of academic publications and results. Use each type of publication as appropriate.
• Know the difference between background literature and related work.
• Do not plagiarize or copy images or text.
• Use proper reference management software when writing your thesis.
APPENDIX
Systematic Mapping Studies and Literature Reviews

**Systematic Mapping Study (SMS)**
- Broad and shallow literature review
- Charts and structures a research area
- Discovers research trends
- Systematic search method, search scope, and criteria for inclusion / exclusion of literature items must be clearly specified
- May be implemented as a combination of automatic analysis (e.g. keyword-based) and manual reviewing

**Systematic Literature Review (SLR)**
- Narrow and deep literature review for a well-defined specific area.
- Built on focused questions to aggregate evidence on a very specific goal
- Quality assessment of primary studies is crucial
  - primary studies without empirical evidence should not be included.

---

