Literature Reviews

Ivo F. Sbalzarini

Outline

- Why?
- How?
- Access to literature
- Finding code
- Using reviews

Caution

It takes time!

Do it at the beginning of your project!
(not nice to find out after the fact
that all has been done before or there was a nice and easy
tool you could have used)

Be organized and keep track!
(Else you end up doing it twice)

Proceed in iterations!

Why?

- Put your work into context and on a academic basis
- Connect your project to the existing knowledge
- Demonstrate your understanding of the state of the art
- Argue for a gap
- Find methods and tools you could use
- Identify potential issues with work you plan to do

How?

- Define your question. Write it down.
- Decide the scope of your review (time, discipline, venues).
- Select the databases and sources.
- Make the searches and **keep track of your results.**
- Read/survey the literature found.
- Categorize / organize / conceptualize the results.
- Write your review.

Define the question

- Start from the broad area (e.g., “compressible Navier-Stokes solvers”)
- Ask yourself questions to focus (e.g., “what Re range?”, “the performance or the accuracy of them?”, …)
- Which aspect of the question are you interested in? (e.g., the numerical methods used? the software engineering aspects? or the flow physics?)
- From the top of your head, or from talking to someone, is there an author, institution, code, … famous in this area?

Decide Scope

- Current or historic? (e.g., the first compressible Navier-Stokes solver ever, or a recent one that is still in use?)
- In what domain? (e.g., solvers for aeronautics? or for biomedical engineering? parallel or sequential solvers?)
- using what methods? (e.g., what programming languages?)
- Type of source material (e.g., conferences, journals, technical reports, source code, only highly cited, etc?)
- Origins (e.g., only solvers from NASA? Only works from the USSR? only open-source works?, etc.)
- Review type (e.g., systematic finding all? exemplary? retrospective?)

Select Sources

- Library catalogues
- Internet sources (Google Scholar, Web of Science, Research Gate, Scopus, ArXiV, BioMedCentral, NCBI, …)
- Books (e.g., amazon, books.com)
- Following citations in other works
- Personal interviews
- Encyclopedias? (usually list facts, but not why and who)
- Wikipedia? (not curated, not peer-reviewed, ever changing, but maybe a good segway)

Make Searches

- Search by keywords
- Keep in mind “synonyms” (e.g., “compressible”=“high Re”, “solver”=“algorithm”, “Navier-Stokes equations”=“momentum conservation”, etc.)
- Use additional fields (date of publication, type of publication) to home in
- Proceed in iterations (e.g., first journal papers of the last decade, then conference papers of the last 5 years, then follow citations back in time to find origins)
- Blend different approaches (e.g., be systematic at first, but then become more exemplary and citation-based).

Make Searches

- Keep in mind different spellings (e.g., “behavior” vs. “behaviour”)
- Keep in mind domain-specific nomenclature (e.g., “portable” vs. “generic” vs. “templated”)
- Keep in mind grammar (e.g., singular vs. plural “solver”/ “solvers”, Capitalization?, dashes?)
- Use wildcards where meaningful (e.g., “solv*” matches “solver”, “solvers”, “solving”, “solvability”, etc.)
- Pay attention to Boolean combinations and correct parentheses (e.g., “Navier-Stokes AND (solv* OR simulat*)”)

Find too much

- Look at the results and check if there is a systematic problem with your searches or if there just really IS a lot of material in this area.
- Refine/sharpen your question if there really is too much in the area originally defined.
- Use more precise search terms or add limits (on date, type, etc.)
- Exclude unrelated areas that use similar terms to reduce false positives (e.g., “solution” in chemistry is not “solution” in computing)

Find too little

- Formulate your question more broadly (e.g., “Navier-Stokes solvers”
- Look for related literature (e.g., “compressible flows”) and see what they cite there and what terms they use to refer to what you are looking for.
- Check if you can find a statement in a conclusions section that says what you are looking for has not been done ever.
- Try different sources (e.g., Google, news archives, youtube) to see if non-academic materials can be found.
- Search for images and videos and follow from there.

Keep Track

- Use a bibliography manager software (e.g., BibTeX, Papers, EndNote, etc.) that works with the text processor you will use to write your review.
- Import all (potential) hits into your database (most online resources support direct dataset export, e.g., Google Scholar exports BibTex records).
- Annotate all your finds with keywords and 2-3 sentences of summary so you know later again what they were.
- Use an indexing system (e.g., “Lastname-firstAuthor:YEAR[a,b,c,…]”) in order to prevent duplicates.
- Record the source and search string used to find a hit.
- attach the PDFs or their DOIs so you can go back to read them.

Read / Survey

- Read along the message tree (reading linearly would take too much time!)
- Keep notes of the top-level messages of the papers read in the annotation field of your bibliography manager.
- Look in particular at the introduction and conclusions and keep notes of: contributions, limitations, open questions.
- Understand WHAT has been done and WHY. Do not care about HOW.
- If a paper raises a question, see if you can find a follow-up where the question has been answered.

Read / Survey

- be critical. Not everything you read is true.
- has the author defined a clear topic / question?
- is there any evident bias (institutional, political, funding, etc.)
- is the discussion credible? Does it outline open questions?
- does it contradict another piece of writing or has it later been disproven?
- are there references you can follow to dig deeper?
- is the topic and the results relevant to your review?

Categorize / Organize

- Is there a classification of the works? (e.g., Reynolds number, parallelism, performance, numerical method, etc.?)
- Are the problem complexes along with the works organize?
- Is there a temporal order or a logical sequence of discoveries / advancements?
- Is there an overarching theme or problem?
- What are the most salient issues dealt with?
- Are there conflicts or arguments in the literature?
- What is clear/accepted, what is open/debated?

Write Review

- Must tell a story and have a top-level message
- Start by stating your original question and discuss it (why this question? how important is it? etc.)
- Formulate your initial thoughts or expectations.
- Provide an overall summary of your search results.
- Present your categorization
- Discuss salient features (e.g., is there an ongoing debate?, is there a clear temporal development?, any particular gaps?)
- State the limits/scope of your search.

Write Review

- Do not simply concatenate all your find in an unstructured succession of bullet items.
- Do not simply describe what you read, but interpret it.
- Discuss the literature found in a logical story following your categorization.
- Summarize views/contributions of multiple authors in relation to your question.
- Highlight links to other problems / disciplines you found without following them deeper.
- Write prose.

Write Review

- Contrast different authors, theories, or methods.
- Evaluate the literature critically. Was anything not credible? did it raise questions without spelling them out? do you believe it? what was the evidence? which projects are still alive?
- End by identifying emerging themes / trends, or gaps
- Close the loop back to your original question and discuss how the literature answers it, or not.

And then?

- Use your review as a basis for an own introduction section.
- Reflect on your own progress in light of your review.
- Use it to apply for fellowships or funds by addressing the gaps identified.
- Reconsider your question and project: is it worth doing / the best you can do given the literature?

Access to Literature

- University libraries (e.g., SLUB) have subscriptions to paid journals. You should be able to get at the literature from within the university network.
- PDFs are usually available only for more recent articles. Older articles can be ordered upon request (2 weeks)
- Historic papers for free: jstor.org
- Many authors post free preprints on their web pages.
- Open-Access journals are on the rise.
- Preprint servers (ArXiV, bioRxiv, ViXrA, etc.)

Anything is OK

- The term “literature” should be broadly understood.
- Papers, books, web pages, blogs, source code, images, videos, user manuals, prototypes, personal communications, etc.
- But pay attention to citing it correctly so others can find it again (DOIs, URIs, etc.)

Find code

- code servers (github, bitbucket, sourceforge, etc.)
- community repositories (ACM TOMS, SIAM, CPC Program Library, NASA, NAG, Numerical Recipes, …)
- github.com/nschloe/awesome-scientific-computing
- Use internet search engines :-)

Find datasets

- wikidata
- kaggle.com/datasets
- github.com/awesomedata/awesome-public-datasets
- data.gov
- community specific (e.g., Broad Image Collection, KEGG, NCBI, …)
- Use internet search engines :-)

Using reviews

- Good and valuable reviews serve to organize the knowledge and put future work on solid grounds.
- They often get published a papers themselves (“review papers”)
- Useful review papers tend to get cited a lot.
- If you find review papers during your search, you often don’t need to go back any further, but simply use that review as a milestone.

Fun

- Can you find the first-ever mention of a given concept (e.g., evolution strategies) regardless the name?
  —> literature archeology.
  (zoom out to superordinate concepts and then branch from there!)
- Can you go into non-english language literature and dig up surprising / unexpected pieces?
- Can you find a piece that contradicts current belief? Can you argue for it?
- Compare your evaluation of a paper, knowing the literature that came after it, with the one given by the authors in their conclusions.