LITERATURE REVIEW

Instructor: Ana Lucia Varbanescu

a.l.varbanescu@uva.nl

Agenda

- A bit of advice on how to approach your literature study
  - Focus on:
    - Goal
    - Method
   - Analysis
    - Writing

- An example

Why literature study?

- To learn the basics of a field

- To grasp the state-of-the-art in a given field

- To understand "gaps" and find new research ideas

In CPP: to extrapolate the knowledge you get in class (fundamentals/basics) to the state-of-the-art analysis

What is the goal ?

We aim to gain (and provide) new knowledge from a set of relevant (and sufficient) publications on a specific topic.
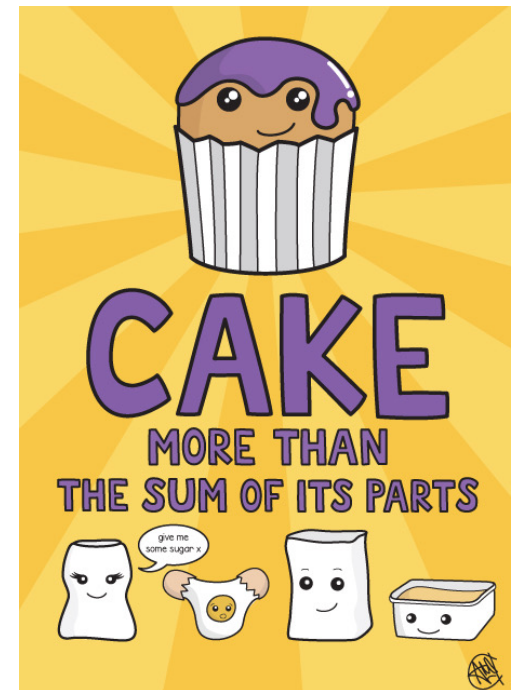
What is the goal?

We aim to gain (and provide) **new knowledge** *from a set of* **relevant (and sufficient) publications** *on a* **specific topic***.*

- Relevant?
  - High quality / timely / …
- Sufficient?
  - Enough to allow for a conclusion
- Sub-domain?
  - Focused domain of interest
- New knowledge?
  - New information, combining facts *across the papers*

What is the "recipe" ?

We aim to gain (and provide) new knowledge from a set of relevant (and sufficient) publications on a specific topic.
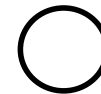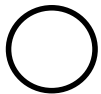
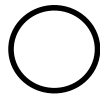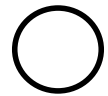*Important steps:*
- Select topic
- Select relevant papers
- Read & reflect
- Analyse
- Put it all together = report

SELECT TOPIC

Not.
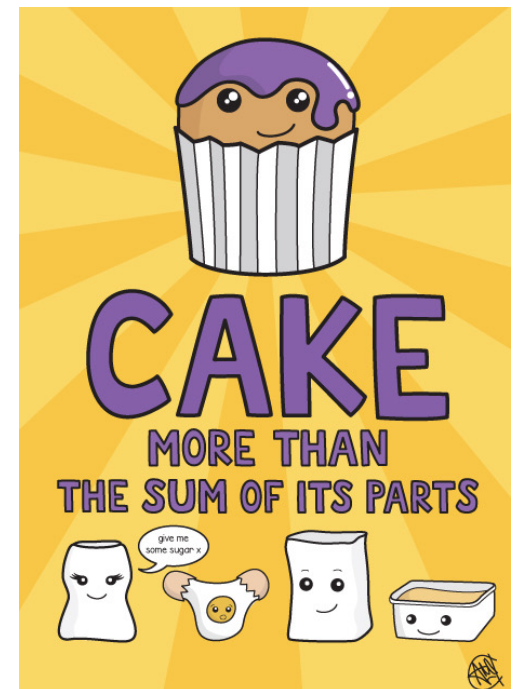
Problem?
Domain is too large, papers are not easily "connected".

Consequence ?
Cannot get sufficient papers for your review.

Better: domain

Problem?
Domain is still too large.

Consequence ?
Papers are very diverse. Difficult to connect and/or compare.
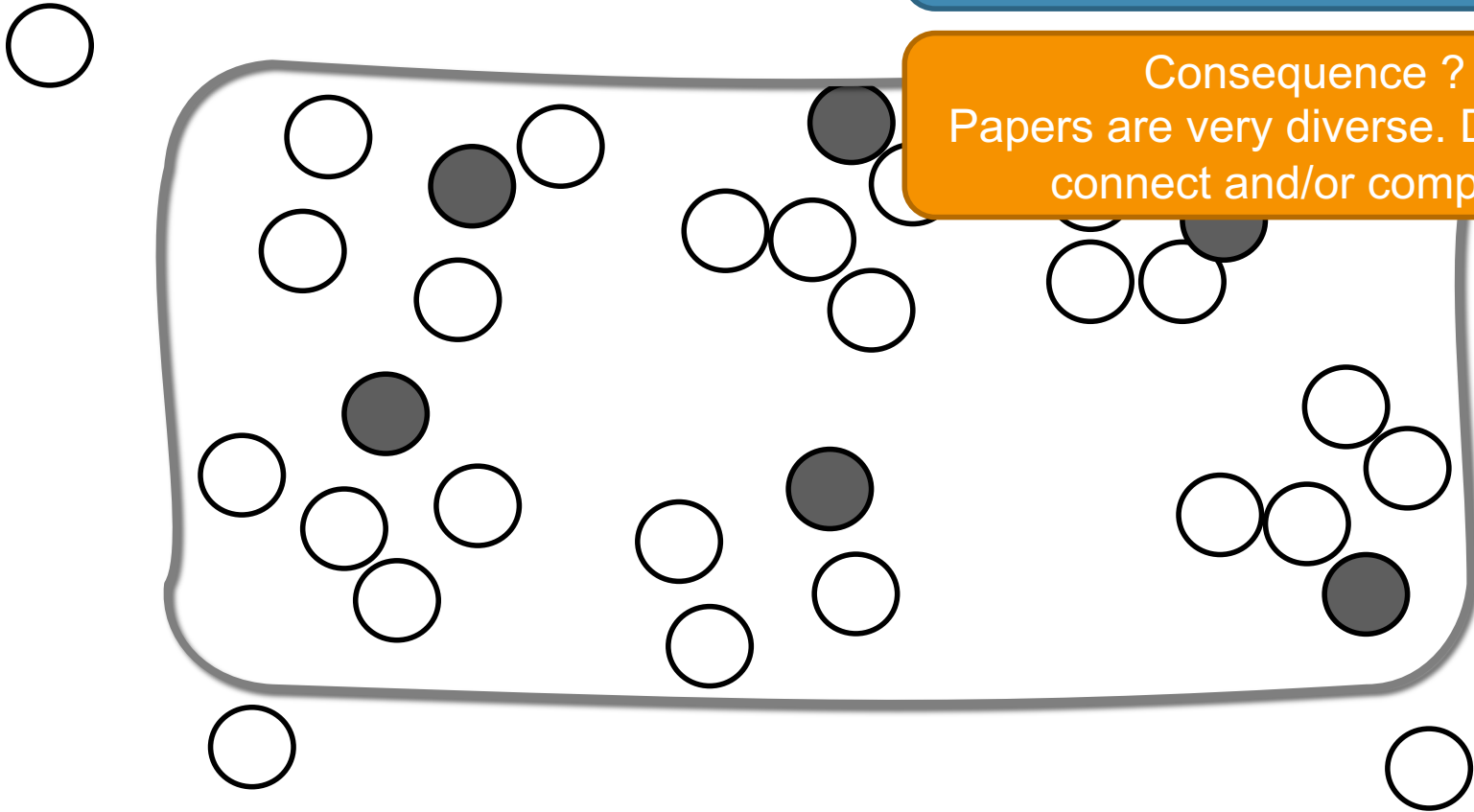
Even Better: cluster, [?]

Class 1

Class 2

Class 3

Not relevant

Getting better:
"Applications running on GPUs"

Advantage?
Domain is still large, but can be clustered!

Consequence ?
Can explain clustering, and focus on one cluster.
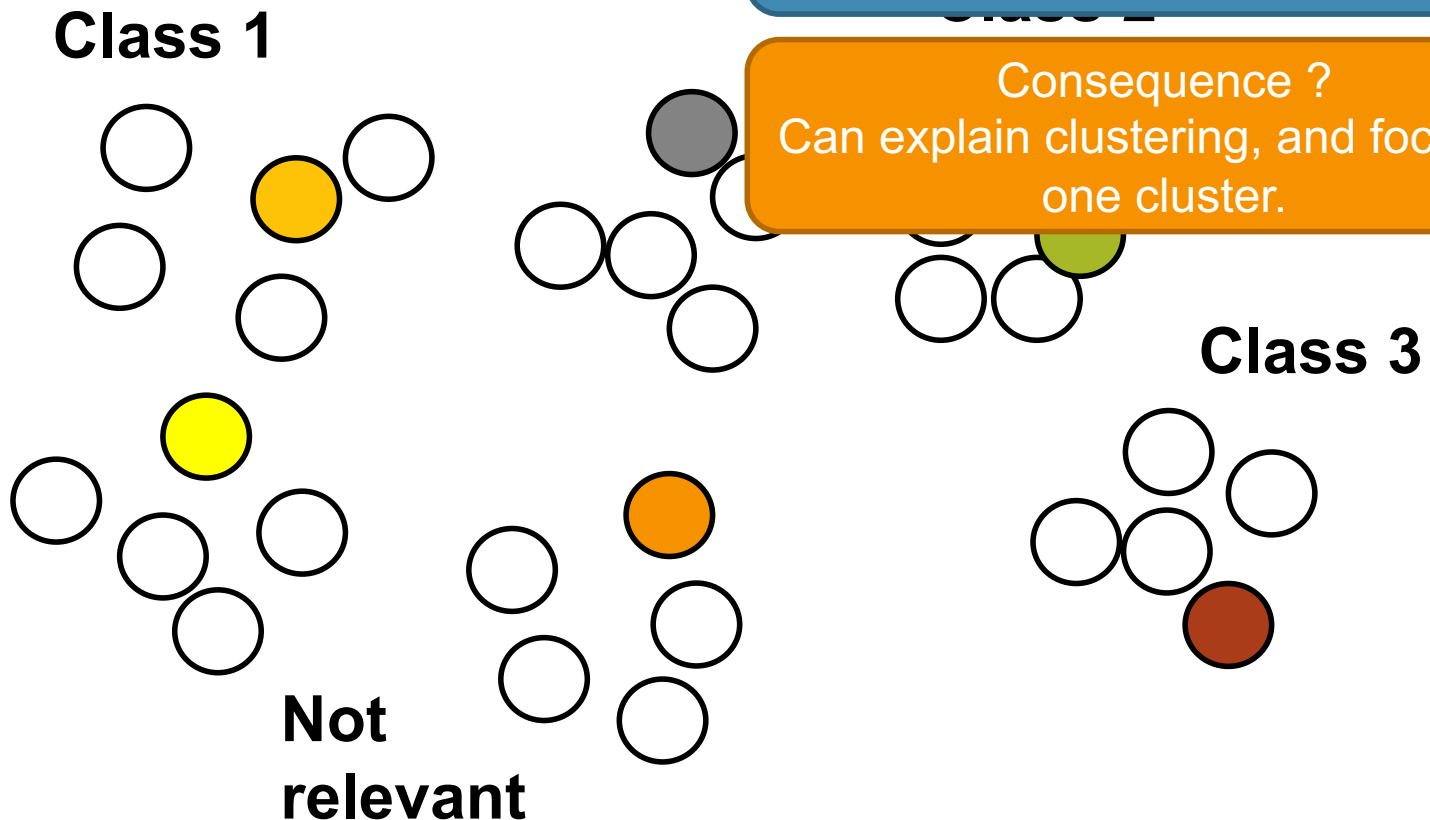
Finally, if possible: a summary…

	Papers	Feature 1	Feature 2
	2	Explain	Explain
Class 2	5	Explain	NA
Class 3	1	Explain	Explain
…	…	--	--

This is your domain now!

… and a research goal/question

- What are the advances in GPUs for gaming ?
- How has performance changed in … ?
- What are the common features of GPUs in … ?
- What are the main features of supercomputers …

How do you define the domain?

You define the domain: what does your review cover?

Potential recipe

1. Start from the given domain
2. Select a part that is of interest to you
   - Because you know it
   - Because you can find it in more accessible sites
     - Blogs, Wikipedia, magazines, …
3. Check for publications
   - Relevant?
   - Sufficient?
4. Repeat …
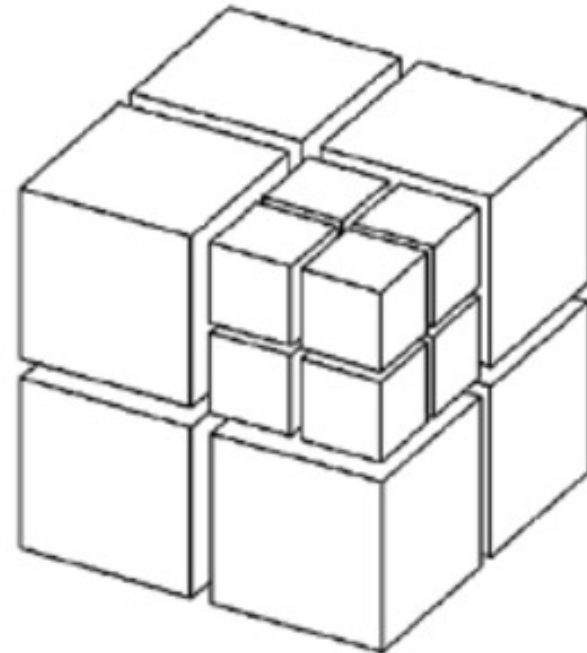
Example: Parallel programming models

- A parallel programming model is a programming tool that has specific primitives for parallelism.
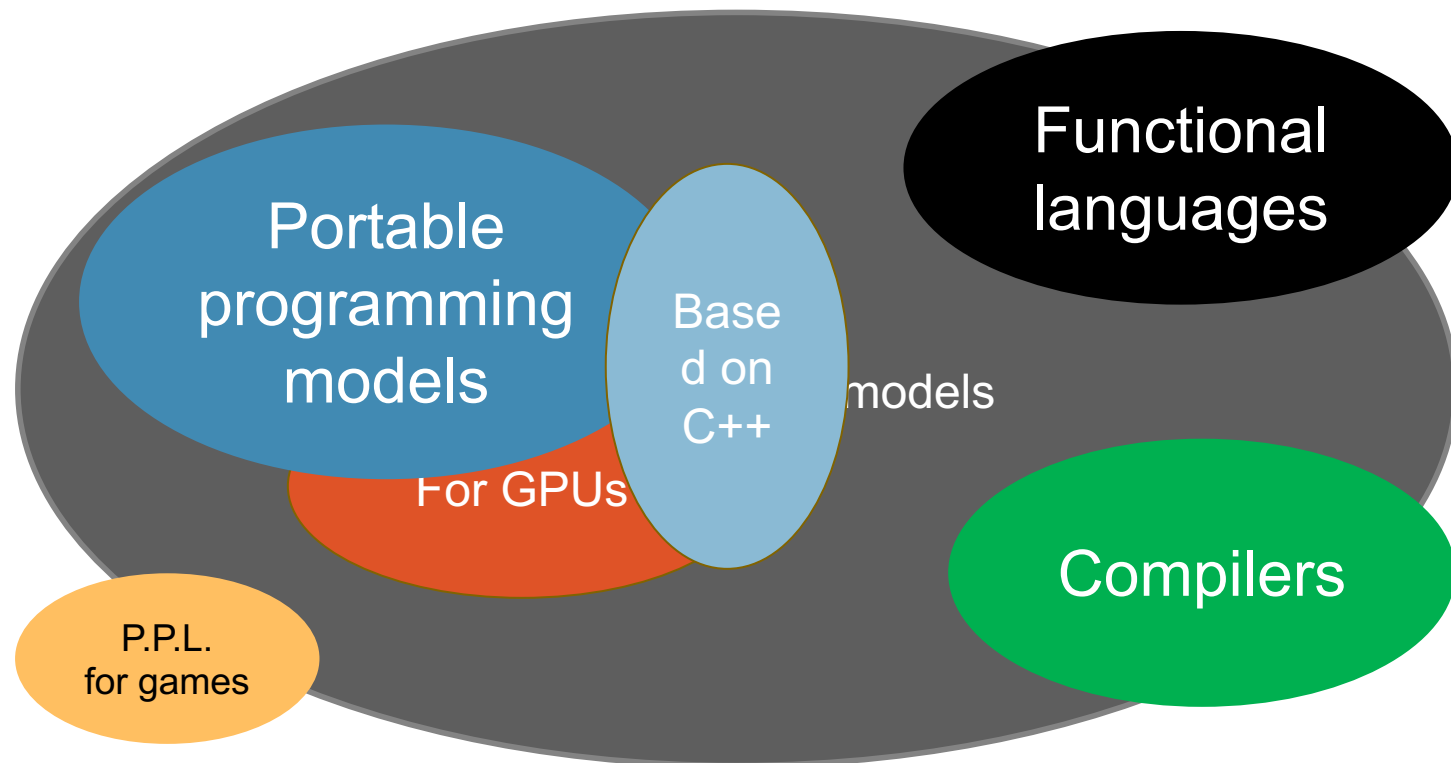
Picking a (Sub-)domain

- Try to ensure:
  - Coherency: the domain is easy to define/explain
  - Size: there are sufficient papers
  - Relevance: there are interesting things to say about it
    - Are you curious about it? Go for it !
  - Comparability
    - Can you find a common "thread" ?

Advice(1): Capture your (sub)domain in the research question and the title…

Advice(2): Start early, ask for advice if needed.

Let's practice

- List of topics
    - Football
    - Bicycles
    - Amdahl's law
    - Supercomputers

- Provide a sub-domain and a research question.

Let's try to do this in groups in 5 minutes!

SELECT PUBLICATIONS

Publications

- Multiple sources to search:
  - Google Scholar
    - Unmoderated
    - Metric-based
  - DBLP
    - Moderated
  - ACM Digital Library/IEEE Explore
    - Moderated, but biased
  - Scopus, WoS
    - Year-based searching, chronologies are easier to find

- Please avoid:
  - Blogs, magazines, Wikipedia as main sources
  - Can be used to delimit your domain

Representative publications

- Must touch on the topic you have.

- … but the results must be ranked:
  - Popularity metrics
    - Paper
      - E.g., Google scholar uses citations
    - Author
  - Objective metrics
    - Age, conference/journal
  - Process metrics
    - Search for X days, read Y papers, …

… but what about surveys?

- If a survey already exists …
  - MUST read and analyze
    - Anything missing ?
    - Anything wrong ?
  - Can select some samples of covered topics for more detailed analysis.
  - Can expand the domain.
  - Can propose a different analysis
    - Any dimension the authors did not cover?

READ AND SUMMARIZE

Read papers…

- DFS (=Depth-First Search)
  - Read every paper to completion
  - Analyze each paper individually first
  - Gather similarities and differences.

- BFS (=Breadth-First Search)
  - Read (multiple) papers in layers
  - Analyze the papers together for similarities
  - Analyze differences per paper

Write about what you read

- For every paper you read: 3-5 paragraphs
    - Summary
    - Positive aspects
    - Negative aspects
    - Specifics for your analysis
    - Relevant related work

    Optional, at times.

- For all papers you read
    - Similarities and/or differences
    - Build/design/define a taxonomy

Your contribution

- The search process.

- The taxonomy and/or analysis
  - Differences and similarities
  - Insights from different papers.

- The papers' analysis.

- Your general(izing) remarks
  - A sketch of the domain
  - Missing issues
  - Future work ideas

ANALYSE / COMPARE

Analysis

Goal: build a classification/taxonomy of the (sub)domain you have selected.

Taxonomy? Classification?

"Systematic arrangement in groups or categories according to established criteria."

- Classification : non-exhaustive
- Taxonomy : exhaustive

- Main challenge: **determine criteria**!

Example: Flynn's taxonomy
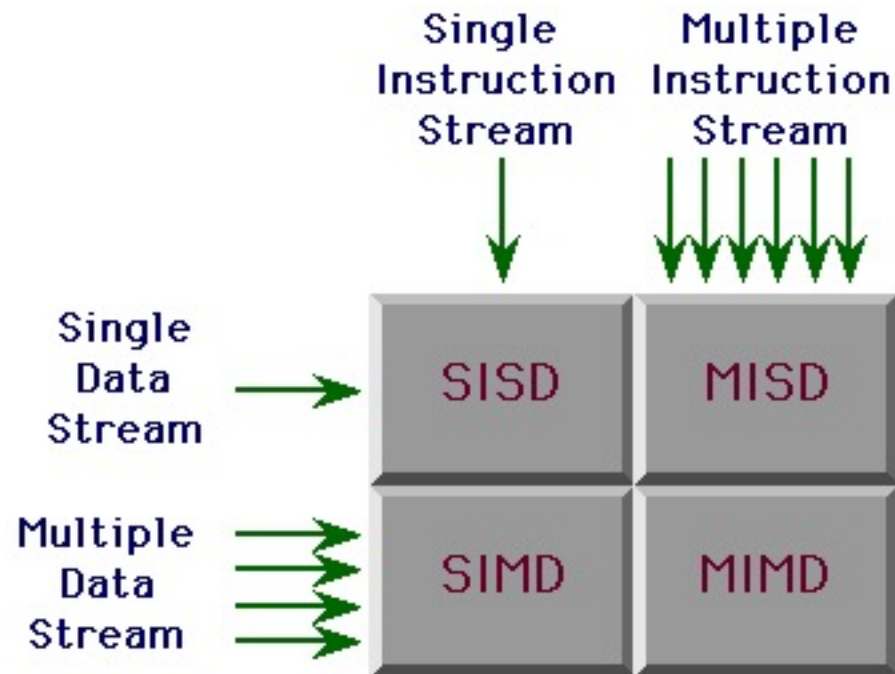
Analysis

Goal: build a classification/taxonomy of the (sub)domain you have selected.

- Bottom-up:
  - Read papers
  - Summarize findings
  - Generalize findings

- Top-down
  - Define a set of expectations/requirements
  - Read papers
  - Discuss coverage of requirements/expectations

Example: bottom-up

*What **are** the common features of GPU algorithms for weather prediction?*

- *Approach:*
  - Read N papers
  - Collect features per paper
  - Compare between papers => defines criteria

- Advantages:
  - start from read papers (so, no "pre-processing")
- Disadvantages:
  - Difficult to find common criteria => difficult to compare …
  - Sparse analysis <= papers have nothing in common ...

Examples: top-down

*What **should be** the main features of GPU algorithms for weather prediction?*

- *Approach:*
  - *Pre-define criteria/features that you WANT to have – "wish-list"*
  - Read N papers
  - Collect features per paper
  - *Compare against the "wish-list"*

- Advantages:
  - Always possible to compare
  - Easier to read the papers
- Disadvantages:
  - Might find that your wish-list is too optimistic/pesimistic
    - Usually we fix this by refining the wish-list

Example: bottom-up

- GPU energy efficiency models:

	Theoretical/Practical	Use	Power savings	Usable across different architectures
Understanding the Future of Energy Efficiency in Multi-ModuleGPUs	Practical	Measuring	n/a	Yes
GPUWattch: enabling energy optimizations in GPGPUs	Practical	Measuring	14.4% / 11.2%	Yes
GPU Energy Consumption Optimization With a Global-Based Neural Network Method	Practical	Power model	4% - 20%	Yes
An integrated GPU power and performance model	Theoretical	Power model	10.99% / 25.85%	No*
Energy efficiency Analysis of GPUs	Theoretical	Power model	n/a	Yes

Example: top-down

- Top500 architectures

Specifications Supercomp	Rank	Architecture	Cores/CPU	Clock speed (GHz)	L1 Caches, size in KB (instruction and data)	L2 Cache size	L3 Cache size
Summit	1	Power9	22	3.07	32; 32	512 KB	120 MB
Sierra	2	Power9	22	3.1	32; 32	512 KB	120 MB
Sunway TaihuLight	3	SW26010	260	1.45	32; 32	256 KB	N.A.
Tianhe-2A	4	Ivy Bridge	12	2.2	32; 32	256 KB	30 MB
Frontera	5	Cascade lake	24	2.7	32; 32	1 MB	38.5MB
Piz Daint	6	Haswell	12	2.6	32; 32	256 KB	30 MB
Trinity	7	Haswell	16	2.3	32; 32	1 MB	40 MB
(ABCI)	8	Skylake	20	2.4	32; 32	1 MB	27.5 MB
SuperMUC-NG	9	Cascade lake	24	3.1	32; 32	1 MB	1.375 MB
Lassen	10	Power9	22	3.1	32; 32	512 KB	120 MB

PUT IT ALL TOGETHER

Step-by-step process

1. Define your (initial) domain (= title)
2. Define approach: bottom-up or top-down
   1. Top-down: define features => sketch taxonomy
3. Refine domain
4. Define search criteria and process
5. Select and read papers
6. Attempt taxonomy/Test taxonomy
   1. Works? Done ☺
   2. Doesn't work -> back to 4, 3, 2.
7. Conclude
   1. Landscape of the (sub-)domain.
   2. Generalized remarks based on taxonomy.
   3. Recommendation for improvement/future work ideas/very interesting points you have found, …

Table of contents [v1]

- Introduction
  - Motivation
  - Definition of the domain / goals
  - Approach: search process, potential search criteria, …
- Representative publications
  - Multiple sections, depending emerging classes.
- A taxonomy/classification/sketch
  - Includes generalizing remarks
  - Could be a table to show differences
- Conclusion/Summary & Future Work
  - Findings
  - Reflection on coverage and better approaches

Table of contents [v1]

- Introduction
  - Motivation
  - Definition of the domain / goals
  - ~~Approach: search process, potential search criteria~~

> … that's not to say these are the real section titles.
> Adapt them to your case / topic / …

  - Includes generalizing remarks
  - Could be a table to show differences
- Conclusion/Summary & Future Work
  - Findings
  - Reflection on coverage and better approaches

Table of contents [v2]

- Introduction
  - Motivation
  - Approach: search process, potential search criteria, …
- Ideal requirements
- Representative publications
  - Possibly multiple sections, depending on how you classify them.
- Coverage of requirements
  - How well does the literature cover the requirements
  - Any paper has them all – highlight
  - Any missing requirement overall – highlight
  - Any observable clustering – highlight
- Conclusion/Summary & Future Work
  - Findings
  - Reflection on coverage and better approaches

Table of contents [v2]

- Introduction
  - Motivation
  - Approach: search process, potential search criteria, …
- Ideal requirements

… that's not to say these are the real section titles.

Adapt them to your case / topic / …

  - How well does the literature cover the requirements
  - Any paper has them all – highlight
  - Any missing requirement overall – highlight
  - Any observable clustering – highlight
- Conclusion/Summary & Future Work
  - Findings
  - Reflection on coverage and better approaches

AN EXAMPLE

SUMMARY

In summary …

- Focus on WHAT you are looking for first.
- Decide on how you will approach things
  - Top-down: you know the field
  - Bottom-up: new field to you
- Define and bound your search
- Put your findings in a taxonomy (table, tree, image, classification …)
- Discuss your findings and highlight things YOU have found
- Summarize/generalize and discuss YOUR ideas for future work
  - In the field or in the analysis. Or both.