AI Master Class, 28/09/2021

Finding literature – and how to read it

Anders Kofod-Petersen
Professor, NTNU
Owner, PiedBoeuf

AI Master Class – recap

- We do computer **science**
  - Science is about method over results
  - Science is about theory over belief
- We need to
  - know what we know
  - be thorough in our approach
  - be able to argue our results
- This is what the AI Master Class is about: you doing the best possible work

Overview of the Master Class, 2021

- ~~14/09/2021 — *Welcome to Dart, introduction* and *how to do research questions*~~
- 28/09/2021 — *Doing structured literature reviews* and *how to read and write a research paper*
- 12/10/2021 — *How to write a thesis*
- 26/10/2021 — *Using HPC at NTNU* and *Reproducibility*
- 09/11/2021 — *How to do qualitative empirical research* (Might change)

https://research.idi.ntnu.no/aimasters/

Your thesis

- You might save the planet
  - However, if you do not know *how* and *why*, and can't describe it — it has little value
- What do you aim for?
  - The average student can **reproduce knowledge**
  - The above average student can **add to knowlede**
  - The good student can **reflect on said addition**
- All of this goes into your thesis!

Method is our friend

- Say this every morning when you look in the mirror: “Method is our friend!”

AI Master Class, 28/09/2021

How to find relevant literature

Anders Kofod-Petersen
Professor, NTNU
Owner, PiedBoeuf

Today's topic

- How do we know what we know?
- Finding literature
  - Structured Literature Review
  - Snowballing
- Reporting

Today's topic

- How do we know what we know?
- Finding literature
  - Structured Literature Review
  - Snowballing
- Reporting

How to do a Structured Literature Review in computer science

Anders Kofod-Petersen

Version 0.2

Contents

How do we know what we know?

- What is you area of research?
- Is it interesting, and why?
- Is it novel, and why?
- Somebody has probably done something similar before
  - What?
  - Who?
  - Why?
  - How?
  - Results?

Your research box

Research questions are the questions that your work should answer

- These are the questions you are evaluated on
- There are the questions your thesis answer
- There are the questions that guide your choice of methods or problem
- They guide your choice of evaluation method, which guides your choice of research questions

Finding literature

- Why put in an effort?
- How do we know what to find?
- How do we know where to look?
- When is a review complete?

What is your approach?

Structured Literature Review (SLR)

- **1 — Planning**
  - ~~Identifying the need~~
  - ~~Commissioning the review~~
  - Specifying the research questions
  - Developing a review protocol
  - Evaluation a review protocol

Structured Literature Review (SLR)

- **1 — Planning**
  - ~~Identifying the need~~
  - ~~Commissioning the review~~
  - Specifying the research questions
  - Developing a review protocol
  - Evaluation a review protocol
- **2 — Conducting**
  - Identification of research
  - Selection of primary studies
  - Study quality assessment
  - Data extraction
  - Data synthesis

Structured Literature Review (SLR)

- **1 — Planning**
  - ~~Identifying the need~~
  - ~~Commissioning the review~~
  - Specifying the research questions
  - Developing a review protocol
  - Evaluation a review protocol
- **2 — Conducting**
  - Identification of research
  - Selection of primary studies
  - Study quality assessment
  - Data extraction
  - Data synthesis
- **3 — Reporting**
  - ~~Specifying dissemination strategy~~
  - Formatting the main report
  - Evaluating the main report

Specifying the research questions

Applied research

- RQ_1: existing solutions to the problem
  - Q_a: different constraints, methods, approaches to RQ_1 solutions
  - Q_b: strength of evidential support
  - Q_c: Implications, when developing your solution

Specifying the research questions

Applied research

- RQ_1: existing solutions to the problem
  - Q_a: different constraints, methods, approaches to RQ_1 solutions
  - Q_b: strength of evidential support
  - Q_c: Implications, when developing your solution

Basic Research

- RQ_1: What are the key areas of investigation for a given technique
  - Q_a: Which areas are interesting and why
- RQ_2: What are the key results in these areas
  - Q_a: strength of evidential support
  - Q_b: Implications, when selecting your area of investigation

Specifying the research questions

Applied research

- RQ_1: existing solutions to the problem
  - Q_a: different constraints, methods, approaches to RQ_1 solutions
  - Q_b: strength of evidential support
  - Q_c: Implications, when developing your solution

Basic Research

- RQ_1: What are the key areas of investigation for a given technique
  - Q_a: Which areas are interesting and why
- RQ_2: What are the key results in these areas
  - Q_a: strength of evidential support
  - Q_b: Implications, when selecting your area of investigation

The review protocol

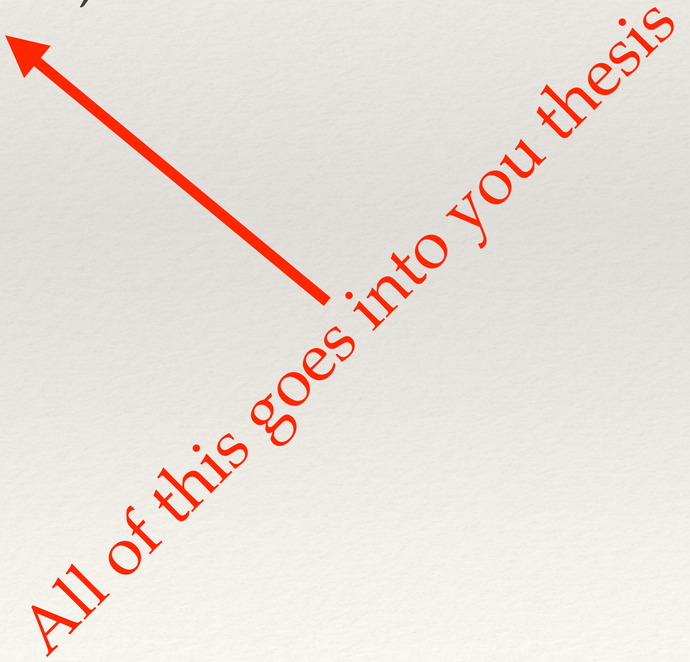
- **This is what makes your research reproducible**
- Search strategy
  - Identification of research (search engines)
  - Relevant terms (search words)
  - Other relevant knowledge (domain experts)
  - Selection criteria
  - Inclusion criteria
  - Quality criteria

Identification of research

- Which sources to be searched
  - ACM, IEEE, ISI, Science Direct, CiteSeer, SpringerLink, Wiley Inter Science, Engineering Village, Google Scholar, etc.
  - Why?
- How to search them?
  - Terms
  - Procedures

Identification of research

- Which sources to be searched
  - ACM, IEEE, ISI, Science Direct, CiteSeer, SpringerLink, Wiley Inter Science, Engineering Village, Google Scholar, etc.
  - Why?
- How to search them?
  - Terms
  - Procedures



What terms are we looking for?

- There are many terms that are more or less synonymous
- We are looking for the the right combinations of the right synonyms
- We are looking for words that have the same semantic meaning (in the domain)
- We might be looking for the right combinations of hypernyms and hyponyms

What terms are we looking for?

- There are many terms that are more or less synonymous
- We are looking for the the right combinations of the right synonyms
- We are looking for words that have the same semantic meaning (in the domain)
- We might be looking for the right combinations of hypernyms and hyponyms

	Group 1	Group 2	Group 3	Group 4
Term 1	$Synonym_1$	$Synonym_2$	$Synonym_3$	$Synonym_4$
Term 2	$Synonym_1$	$Synonym_2$	$Synonym_3$	$Synonym_4$
Term 3		$Synonym_2$	$Synonym_3$	
Term 4			$Synonym_3$	
Term 5			$Synonym_3$	

What terms are we looking for?

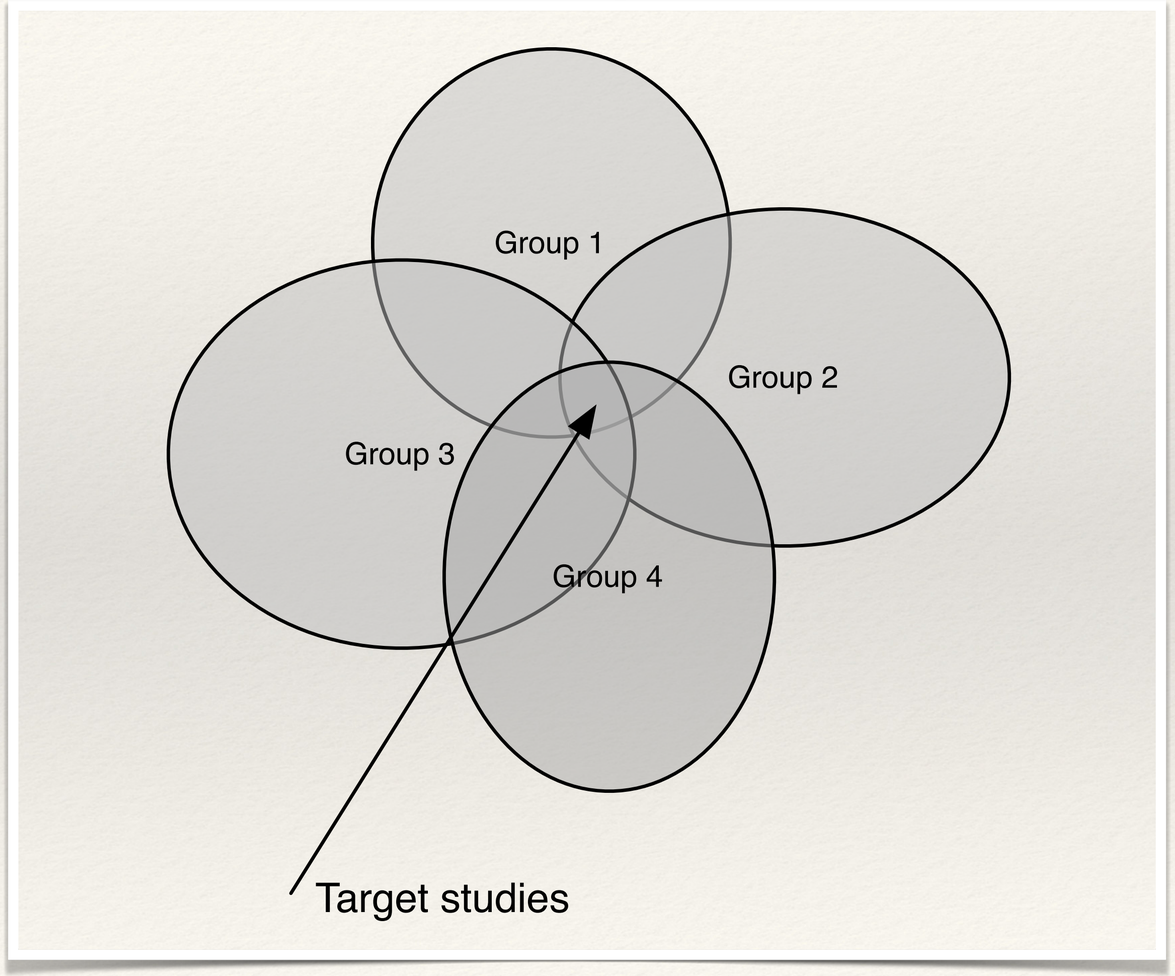
Search strategy is done by implementing AND, and OR between terms:

([G1, T1] OR [G1,T2]) AND

([G2,T1] OR [G2,T2] OR [G2,T3]) AND

([G3,T1] OR [G3,T2] OR [G3,T3] OR [G3,T4] OR [G3,T5]) AND

([G4,T1] OR [G4,T2)



Selection of primary studies

- Searching will generally return far too many results. We can limit the set by:
  - Remove duplicates (keep the highest ranking source)
  - Remove the same study published in different sources (keep the highest ranking source)
  - Remove studies published before a certain date (or even after)

Quality assessment

- Abstract inclusion screening
- Full text inclusion screening
- Full text quality screening

Criteria identification	Criteria
IC 1	The study's main concern is $\mathcal{P}$
IC 2	The study is a primary study presenting empirical results
IC 4	The study focuses on $\mathcal{C}$
IC 5	The study describes an $\mathcal{S}$
QC 1	There is a clear statement of the aim of the research
QC 2	The study is put into context of other studies and research

Final quality assessment

- QC 1 Is there is a clear statement of the aim of the research?
- QC 2 Is the study is put into context of other studies and research?
- QC 3 Are system or algorithmic design decisions justified?
- QC 4 Is the test data set reproducible?
- QC 5 Is the study algorithm reproducible?
- QC 6 Is the experimental procedure thoroughly explained and reproducible?
- QC 7 Is it clearly stated in the study which other algorithms the study's algorithm(s) have been compared with?
- QC 8 Are the performance metrics used in the study explained and justified?
- QC 9 Are the test results thoroughly analysed?
- QC 10 Does the test evidence support the findings presented?

Final quality assessment

- QC 1 Is there is a clear statement of the aim of the research?
- QC 2 Is the study is put into context of other studies and research?
- QC 3 Are system or algorithmic design decisions justified?
- QC 4 Is the test data set reproducible?
- QC 5 Is the study algorithm reproducible?
- QC 6 Is the experimental procedure thoroughly explained and reproducible?
- QC 7 Is it clearly stated in the study which other algorithms the study's algorithm(s) have been compared with?
- QC 8 Are the performance metrics used in the study explained and justified?
- QC 9 Are the test results thoroughly analysed?
- QC 10 Does the test evidence support the findings presented?

We will get to writing it up

Snowballing

- Start with a few high quality / high impact papers (start set)
- Now roll the snowball. Either backwards or forward
- As always, do things methodically. It should be reproducible

Backward snowballing

- Set up explicit inclusion and quality criteria
- Repeat until no more papers found
    1. Go through the reference list in new papers in the start set and exclude papers that does not conform to the inclusion criteria
    2. Remove papers already found
    3. Add found papers to your start set
- Apply quality criteria
- Write the synthesis

Forward snowballing

- Set up explicit inclusion and quality criteria
- Repeat until no more papers found
    1. Go through each new paper a figure out who have cited this paper and exclude those that does not conform to the inclusion criteria
    2. Remove papers already found
    3. Add found papers to your start set
- Apply quality criteria
- Write the synthesis

Writing up your related work

- The papers found of sufficiently high quality constitutes your related work
- Related work is often in addition to background
- Your related work is where you demonstrate what kind of student you are
  - The average student can **reproduce knowledge**
  - The above average student can **add to knowlede**
  - The good student can **reflect on said addition**

Am I reproducing, adding or reflecting?

“That is quite different than what I will be dealing with I find the paper interesting”

Am I reproducing, adding or reflecting?

"That is quite different than what I will be dealing with I find the paper interesting"

"This is a problem I will be facing and I think this papers approach is interesting"

Am I reproducing, adding or reflecting?

"That is quite different than what I will be dealing with I find the paper interesting"

"This is a problem I will be facing and I think this papers approach is interesting"

"I think this paper is interesting since it demonstrates…"

Am I reproducing, adding or reflecting?

“That is quite different than what I will be dealing with I find the paper interesting”

”Their architecture has several changes compared to…”

“This is a problem I will be facing and I think this papers approach is interesting”

“I think this paper is interesting since it demonstrates…”

Am I reproducing, adding or reflecting?

"That is quite different than what I will be dealing with I find the paper interesting"

"This is a problem I will be facing and I think this papers approach is interesting"

"Their architecture has several changes compared to…"

"I think this paper is interesting since it demonstrates…"

"The version of the xx problem that is dealt with in this paper is almost the same as the problem I intend to take on in my paper"

Am I reproducing, adding or reflecting?

“That is quite different than what I will be dealing with I find the paper interesting”

”Their architecture has several changes compared to…”

“The version of the xx problem that is dealt with in this paper is almost the same as the problem I intend to take on in my paper”

“This is a problem I will be facing and I think this papers approach is interesting”

“I think this paper is interesting since it demonstrates…”

“XXX did apply YYY method to problem ZZZ”

Am I reproducing, adding or reflecting?

- State of the Art ≠ List of Contributions
  - ! Short summary of paper
  - ! Why the author liked it or not
- What is done ≠ Why something is done
  - This should be the opinion of you in the context of the State of the Art

Am I reproducing, adding or reflecting?

- State of the Art ≠ List of Contributions
  - ! Short summary of paper
  - ! Why the author liked it or not
- What is done ≠ Why something is done
  - This should be the opinion of you in the context of the State of the Art

Applied research

- RQ1: existing solutions to the problem
  - Qa: different constraints, methods, approaches to RQ1 solutions
  - Qb: strength of evidential support
  - Qc: Implications, when developing your solution

Basic Research

- RQ1: What are the key areas of investigation for a given technique
  - Qa: Which areas are interesting and why
- RQ2: What are the key results in these areas
  - Qa: strength of evidential support
  - Qb: Implications, when selecting your area of investigation

Why?

- Say this every morning when you look in the mirror:
  1. “Method is our friend!”
  2. “Why!”

AI Master Class, 28/09/2021

What's the deal with papers?

Anders Kofod-Petersen
Professor, NTNU
Owner, PiedBoeuf

Why do we want to read a paper?

- Somebody has already done some similar work
- It is current (bleeding edge research)
- It is reproducible
- It has raw data
- It shows logic
- It informs our own research

Why do we want to write a paper?

- Knowledge has no value in my head
- We wish to report on the state of the art
- We wish to have our research scrutinised
- We wish for the recognition of our peers
- We need research grants

What constitutes a good paper

- QC 1 Is there is a clear statement of the aim of the research?
- QC 2 Is the study is put into context of other studies and research?
- QC 3 Are system or algorithmic design decisions justified?
- QC 4 Is the test data set reproducible?
- QC 5 Is the study algorithm reproducible?
- QC 6 Is the experimental procedure thoroughly explained and reproducible?
- QC 7 Is it clearly stated in the study which other algorithms the study's algorithm(s) have been compared with?
- QC 8 Are the performance metrics used in the study explained and justified?
- QC 9 Are the test results thoroughly analysed?
- QC 10 Does the test evidence support the findings presented?

How to write a paper?

1. Do the research
2. Make sure your method is solid
3. Figure out who your audience is
4. Find a publication channel
5. Write it up
6. Submit
7. Read reviews
8. Adjust the paper
9. Repeat 6-8 until successful
10. Celebrate

How to write a paper?

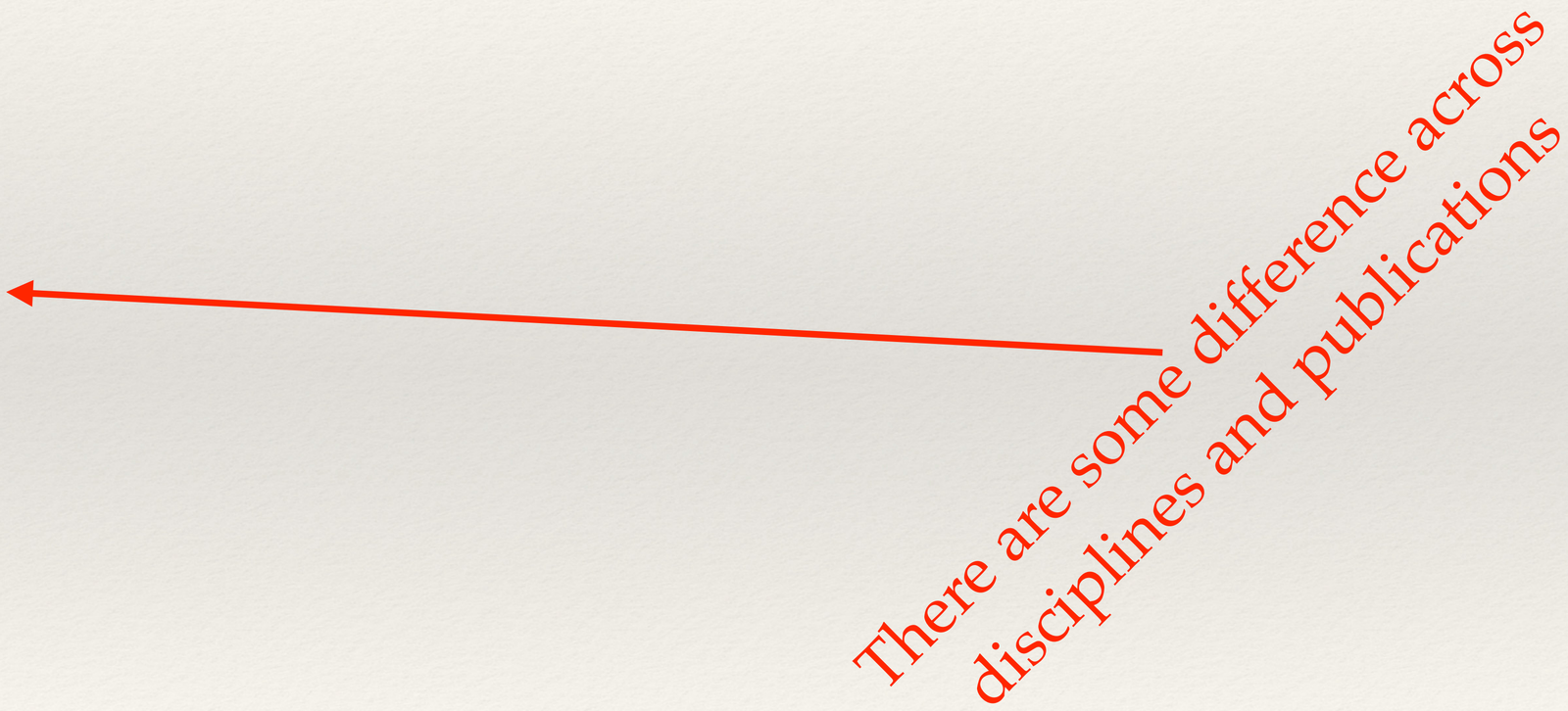
1. Do the research
2. Make sure your method is solid
3. Figure out who your audience is
4. Find a publication channel
5. Write it up
6. Submit
7. Read reviews
8. Adjust the paper
9. Repeat 6-8 until successful
10. Celebrate

Structure of a scientific paper

1. Title
2. Abstract
3. Key words
4. Main text (IMRAD)
    1. Introduction
    2. Methods
    3. Results
    4. Discussion
5. Conclusion
6. Acknowledgements
7. References
8. Supplementary data

Structure of a scientific paper

1. Title
2. Abstract
3. Key words
4. Main text (IMRAD)
    1. Introduction
    2. Methods
    3. Results
    4. Discussion
5. Conclusion
6. Acknowledgements
7. References
8. Supplementary data



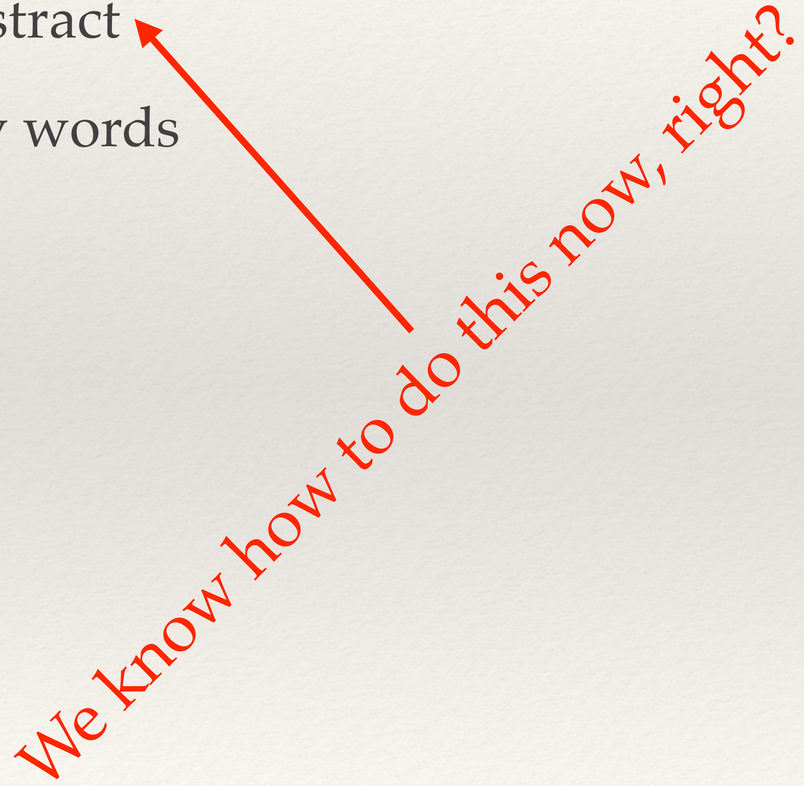
(A)IMRAD

1. Title
2. Abstract
3. Key words

1. Main text (IMRAD)
2. Introduction
3. Methods
4. Results
5. and
6. Discussion

(A)IMRAD

1. Title
2. Abstract
3. Key words



1. Main text (IMRAD)
2. Introduction
3. Methods
4. Results
5. and
6. Discussion

(A)IMRAD

1. Title
2. Abstract
3. Key words

We know how to do this now, right?

1. Main text (IMRAD)
2. Introduction
3. Methods
4. Results
5. and
6. Discussion

We might want to add a *related work*

Introduction

- Why do the authors do this work?
- What is the research question?
- What is the tested hypothesis?
- What is the purpose of this research?

Introduction

- Why do the authors do this work?
- What is the research question?
- What is the tested hypothesis?
- What is the purpose of this research?

Related work

- Which shoulders are the authors standing on (paraphrased)?
- Key references supporting background information
- Refer to the authors' previous preliminary work (if any)
- Refer to the authors' closely related work (if any)

Related work

- Which shoulders are the authors standing on (paraphrased)?
- Key references supporting background information
- Refer to the authors' previous preliminary work (if any)
- Refer to the authors' closely related work (if any)

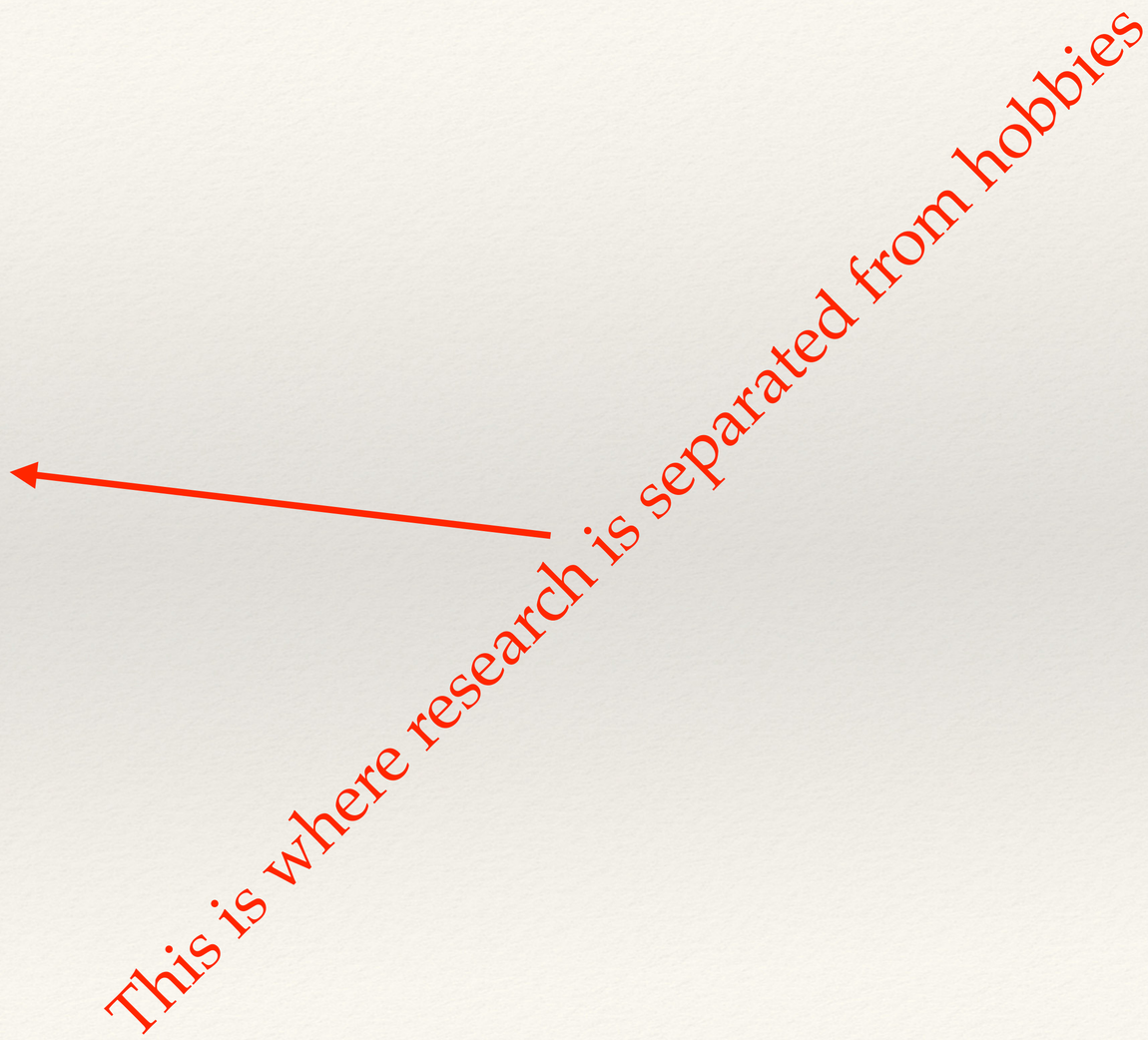
Beware of dragons — here be self-reference

Methods

- Why do the authors do this work?
- What is the research question?
- What is the tested hypothesis?
- What is the purpose of this research?

Methods

- Why do the authors do this work?
- What is the research question?
- What is the tested hypothesis?
- What is the purpose of this research?

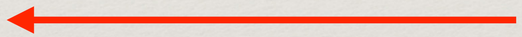


Results

- What answers were found to the research question
- What did the study find
- Was the tested hypothesis correct

Results

- What answers were found to the research question
- What did the study find
- Was the tested hypothesis correct



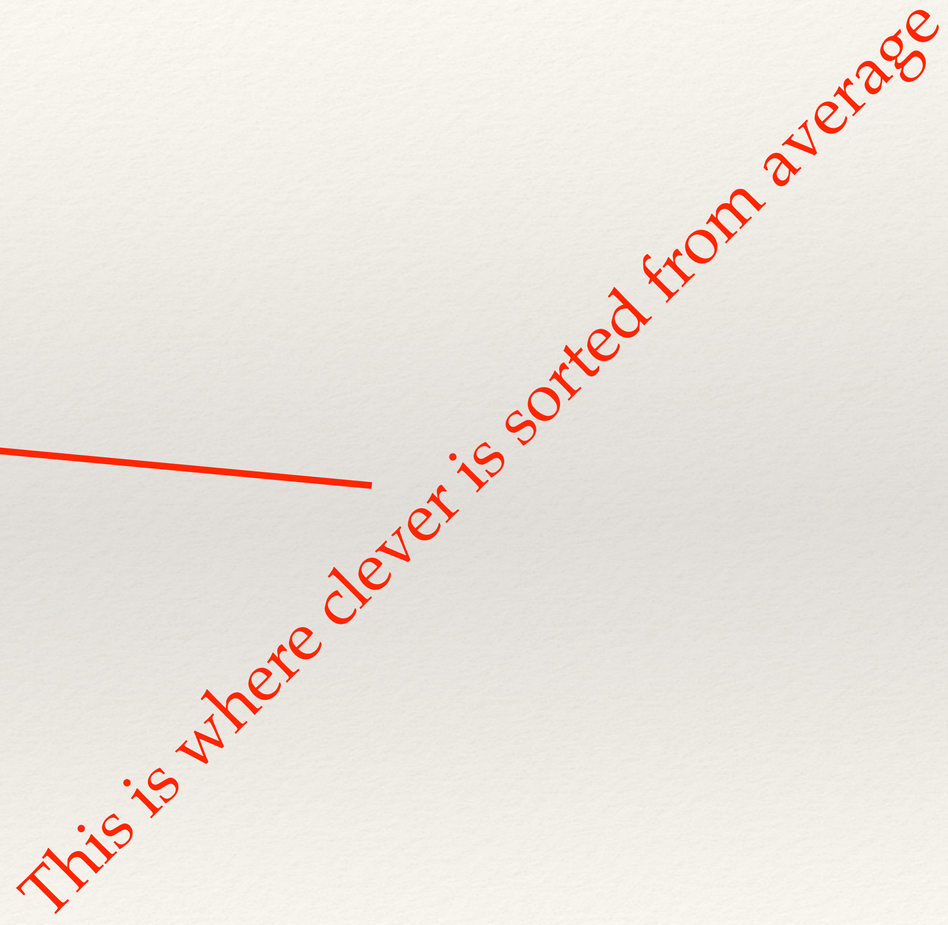
Results! Why, man, I have gotten a lot of results!
I know several thousand things that won't work.
-- Thomas Edison

Discussion

- What might the answer imply
- Does it matter (outside this little box)?
- How does it fit in with related research?
- Are the any threats to validity?

Discussion

- What might the answer imply
- Does it matter (outside this little box)?
- How does it fit in with related research?
- Are the any threats to validity?



Conclusion

- This is not a new abstract
- It is not a listing of results
- It is a clear scientific justification of the work
- Suggest future research

Conclusion

- This is not a new abstract
- It is not a listing of results
- It is a clear scientific justification of the work
- Suggest future research

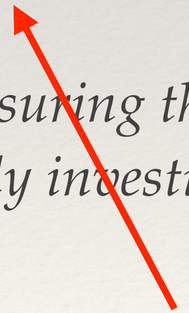
This is short – we already read the paper!

Acknowledgements

- Thank people who have contributed (but not enough to be a co-author)
- What is an author:
    1. *Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND*
    2. *Drafting the work or revising it critically for important intellectual content; AND*
    3. *Final approval of the version to be published; AND*
    4. *Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.*
- Everybody else goes here.

Acknowledgements

- Thank people who have contributed (but not enough to be a co-author)
- What is an author:
    1. *Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND*
    2. *Drafting the work or revising it critically for important intellectual content; AND*
    3. *Final approval of the version to be published; AND*
    4. *Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.*
- Everybody else goes here.



The Vancouver Recommendations

References

- Just follow the guidelines for the channel
- Don’t do too many self reference!

Supplementary data

Tips and tricks

- Read a lot of papers
- Write a lot
- Use LaTeX!
- Get somebody to read review
- Don’t do language cock-ups
- Write clear and concise
- Adhere to writing style
- …

Overview of the Master Class, 2021

- ~~14/09/2021 — *Welcome to Dart, introduction* and *how to do research questions*~~
- ~~28/09/2021 — *Doing structured literature reviews* and *how to read and write a research paper*~~
- 12/10/2021 — *How to write a thesis*
- 26/10/2021 — *Using HPC at NTNU* and *Reproducibility*
- 09/11/2021 — *How to do qualitative empirical research* (Might change)