



Information Technology

Introduction to Candidature

(including the formulation of research questions)



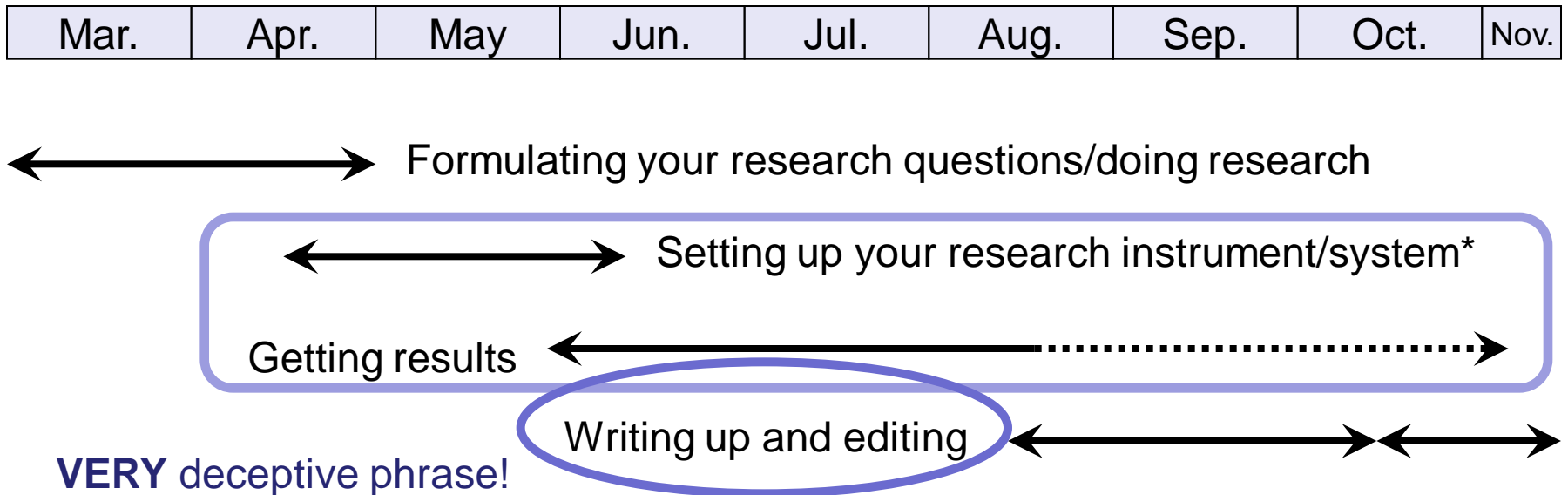
What will be discussed

- Time management issues specific to honours students
- Research topics vs. research questions

Writing in the honours year

- An honours year is not a year but 8 months
- Aside from all the coursework and laboratory work, you still have to spend some time on writing related work, which includes:
 - The formulation of research questions
 - Researching existing literature
 - Drafting
 - Editing and formatting

Writing a thesis (a sample timeline)



* This often overlaps significantly with the 'Getting the results' phase

What you have to consider when planning to write a thesis

- 1 to 2 months to formulate your research questions
- 2 months for drafting (not once but few times)
- 1 month for editing and formatting

What you will be doing (for the next one or two months)

- You will most likely be working on your research questions
- **Research topics** and **research questions** are different
- The difference between the two will be the first hurdle that you have to overcome...
 - ... even if you already know what you actually have to do technically/practically for your project

Research topics and research questions

- Research topics
 - Evolution (C. Darwin)
 - DNA (J. Crick and F. Watson)
 - Ontology (K. Gödel)
- Research questions
 - Do species evolve from another successively?
 - What does a DNA structure look like?
 - Can we prove the existence of God?

Research topics and research questions in IT

- **Research topics:** What your thesis is about
- **Research questions:** What your thesis actually deals with
 - **Research topic:** Intrusion detection in WSN
 - **Research question:** Does a Graph Neuron approach work well in the context of WSN intrusion detection? (from Alex's honours project)

Research topics and research questions in IT

- Research topic and questions decide the nature of your research (methodology, approach etc.) as well

What the thesis is about

- **Research topics:** Intrusion detection in WSN

The question to which the thesis responds

- **Research questions:** Does a Graph Neuron approach work well in the context of WSN intrusion detection?

What the thesis actually is

- *This research is **an experimental study** to assess the capability of GN in the WSN environment.*

From last year's topics...

- **Simulation in Geostatistics**

← **Topic**

- Consider the problem of determining the amount of soil contamination given several samples in a contaminated site. If you also have a geostatistic model of how the contamination varies in the site, then you could use simulation to estimate the distribution of the soil contamination. In this project we will investigate various methods of doing simulation in geostatistics.

↑
Potential question

Formulating research questions

- With the sentence “In this project we will investigate various methods of doing simulation in geostatistics”, you can formulate **research questions** like:
 - What kind of simulation techniques are available?
 - Are they feasible in different environments?
 - How can an appropriate technique be selected?

Understand what your thesis actually is

- With the formulated **research questions**, you can actually see what to do to answer/respond the questions
- And these will also tell you **what the project is**. In the example, the thesis will likely be...
 - An **assessment** of available techniques in geostatistical simulation
- **And this is what you have to do early in your research project**

Another example from this year

Bayesian Evolutionary Trees (12 or 24-pts)

Given K DNA sequences, or protein sequences, the multiple-alignment problem and the evolutionary-tree problem are "chicken and egg" problems: Given an optimal evolutionary-tree one can search for an optimal multiple-alignment of the K sequences (and we know how to do that [LW94]), and given an optimal multiple-alignment of the K sequences one can search for an optimal evolutionary-tree.

An evolutionary-tree is a special case of a Bayesian net [KN04]: a tree rather than a general DAG, with discrete variables (over $\{A,C,G,T\}$ if DNA) at each node, known values at all leaves, and missing values at all internal (ancestral) nodes, e.g., see fig.4 [LW94].

The project is to modify the CaMML algorithm for Bayesian nets to infer evolutionary trees given a multiple alignment. If that is successful, the feasibility of simultaneously solving the multiple-alignment and evolutionary-tree problems will be investigated.

The project involves algorithms, probability and machine learning. You should have good results in mathematics, algorithms and data structures, and formal methods, and must have discussed the project with LA or KK before submitting "the form".

Formulating secondary research questions

- Once you can clarify exactly what it is that your research will address (i.e. **a primary research question**), you can use this to formulate **secondary questions...**
- These will be any questions that will need to answered in order to address your primary research question, and they can be formulated by considering the component parts of your research

Formulating secondary research questions

- For example...
 - Is there any information that will need to be determined before you can answer your primary research question?
 - Are there any impacting factors that will need to be considered?
- A hierarchical structure of questions will allow you to relate everything back to a single focus (namely, your primary research question)



Drawing upon these questions as you plan your research tasks

- These secondary questions may influence the planning of your research tasks, so that your research can accommodate the answering of these secondary questions
(with the ultimate aim of answering your primary research question)



Summary of main points

- Formulate your research questions as soon as possible
 - Understand what it is that your thesis will be
- Keep writing, but know exactly what you are writing
 - Always keep a perspective both in terms of the candidature, and thesis writing