How to write a scientific essay

Olaf Wolkenhauer

Dept of Systems Biology & Bioinformatics, University of Rostock

Web address: **www.sbi.uni-rostock.de** Twitter: **@OlafWolkenhauer**

You have been assigned a five page essay, focussing on a topic related to what the Department of Systems Biology & Bioinformatics is interested in. You are wondering what you should do and how?

Welcome to the Department of Systems Biology & Bioinformatics. If the assignment was not your choice, we are sorry about that and hope we find a way to make your effort and our time worthwhile.

What about?

The first question is probably "What should I write about?" The most important part of an answer to this question is to find a question that you are actually interested in. If you just copy and paste things from the Internet to hand in something, you will not have learned anything, and neither you, or we will be happy with the result. Therefore, try to identify a question that raises some interest with you!

Our Department is interested in mathematical modelling and data analysis, and we deal with everything from theoretical mathematical research, to the development of software tools. We have basic research projects, and various applied projects, including industry collaborations.

You can ask yourself a question related to mathematical models using category theory, differential equations, game theory, graph theory, stochastic models … You can ask yourself a question about machine learning models for classification, deep learning models for image analysis, statistical models for decision making … it should be possible to find something of interest in these broad areas.

You can discuss the mathematics of a deep learning algorithm, or you discuss an ethical dilemma linked to the use of deep learning algorithms in autonomous cars, or in medicine. The choice is yours.

Rather than writing an essay, you may want to just read one publication and review it? Or, read a review and summarise it? Fine with us! The "what you write about" is less important than the "how you write".

Rather than reviewing something or reading up on a topic, you want to do something yourself, implement or program something? Go for it and use the text to document your experience. Given that you probably have not too much time for this, the goal is not to produce some amazing result. Instead, you may describe the workflow, stumbling blocks, things that could have been easier, or describe what surprised you. Write the essay as if you want to share your experience and insights with a friend.

How to go about it?

The best way to start writing is NOT to write full sentences. You start writing an Outline, that is, headings with short, one line bullet points below. These can be more easily be shuffled around until the structure of your essay becomes clear.

Each bullet point under a section heading could be written out as a paragraph. Each paragraph should make one and only one point. The first sentence of a paragraph sets the context, and the last sentence of the paragraph is the punch line, bring the take away message.

The Internet is full of sites with tips on scientific writing and depending on what format you want to pursue (review, summary of a review, analysis etc) there are different ways to structure your text. Here is one example for how to go about writing a scientific essay:

• Introduce the topic to the reader and explain what you are trying to achieve.
• Review the relevant literature and introduce different approaches. If you feel confident, don't shy away to make personal statements about your judgement of the issues.
• Present data and results, objectively and factually.
• Identify any weaknesses, discuss challenges and opportunities.
• Conclude with a summary and conclusions.
• Provide references for your text, ensure you cite them properly.

Reviewing a scientific paper, or an approach, the following questions can help you summarise the work:

• What is the context?
  – What in general?
  – What in particular?
• What is the question?
  – What problem is addressed?
  – What is the research gap?
• What has been done?
  – What approach is taken?
  – Which technology/methodology is used?
• What has come out?
  – What is the main result? (expected outcome)
  – What is the advance? (difference to other works)
• What's the value?
  – What does this contribute to?
  – How does this add to previous knowledge?

You were told to write five pages, but this is really an upper limit. More is not better and the challenge is usually to keep things concise. In project meetings, progress reports, at conference talks and for a project pitch, you have often only 5 to 15 minutes to present your work. An abstract, or project summary is typically half a page. If you require less than five pages, don't worry, this can be perfectly fine. If you use a diagram or you create a table to summarise something, this can be worth several pages of text.

Most important in writing is to make every sentence count. Question every sentence, whether it could be shortened or whether it is really necessary. This is why the top tip is to start with an outline that has just headings and bullet point lists below. The bullet points are keyword or hints to what you want to communicate under the heading. Only after this has been finished, you expand the bullet points into sentences and paragraphs!

Why?

You may be forced into this assignment and it is a good idea to make the most of it. If you just copy and paste texts from the Internet together, you will not learn anything. Remember, this exercise is not just about a scientific topic, or understanding some scientific concept; the assignment is also about learning scientific writing.

As a student you learn how to solve problems, how to produce results. What we usually do not learn at University is how to communicate our work. This is a problem because experience shows that the

communication of results is incredibly important. Whether you continue in academia, or industry, you will have to be able to present your work in say 200 words, in 15 min, effectively and convincingly. It does not matter whether one person or 100 people have worked on the project, whether it took one week, or five years - it must always be possible to communicate a project in a very concise form. This is a skill that requires practice, a lot of practice and you better start now working on this.

To some extend you only as clever as others believe you are. Formulating an interesting question, analysing evidence, summarising literature, drawing conclusions and making this interesting and convincing, isn't easy. This assignment can be a useful opportunity to practice and learn something new.

Make sure you learn something with this exercise, and make sure you have reasons to enjoy it. If the choice of your own question is not enough, consider using the assignment to try something new. For example, try using a new tool for writing, like Overleaf, using a mindmap tool to create an outline, or try using using Markdown. At some point we all have to use a drawing tool to create diagrams (often in order to summarise things, like workflows).

It has become quite common for scientists and engineers to communicate through Twitter and blogs. Sharing code, workflows, notebooks and reviews is not only a contribution towards transparency and reproducibility. Sharing is also caring. Writing a blog maybe used to promote opinions but it could also be a way to contribute to a community, to help others. Consider sharing an experience with the assignment.

What for?

You will not be given a mark for your essay but a decision must be made whether you passed. You may therefore ask "How is it decided whether I pass?"

A 60% mark is often used to characterise a work that has both strong and weak features with the stronger features being in the majority. In the present case, we take a 60% mark as the threshold. Less evidence for critical awareness and independent thinking is typically a reason why not more than 80% was achieved. 80% corresponds to work that is mainly of professional standard, but has few shortfalls. Displays good understanding of the subject within a wider context. A mark of 80% is usually in indication that the author has reached beyond the essential material.

What are the criteria that contribute towards a 60% mark?

- The author demonstrates background understanding, has conducted a literature search, or has done some exploratory work.
- The difficulty of the question, embracing new ideas, and originality.
- Appropriate choice of method(s), clarity and accuracy, adequate acknowledgement of sources.
- Clear structure, clearly expressed aim, (general) goals and (specific) objectives.
- Clarity of text, grammar, spelling, casual mistakes (typos), ease of reading/understanding.
- Use of diagrams and tables, their legibility, visual clarity of information, labelings.

Judging the work of others, giving marks is the worst part of my job. I thoroughly dislike it. But I also dislike cheating and people who try to get away without making an effort. Please consider this assignment not a chore or something you are forced to do against your will. Please try to make it a useful exercise for you by answering a question you are interested in, or by developing your writing and presentation skills.

Still not sure?

Ok, I tried but if you still would prefer hints to what to discuss, here are some ideas. All of our projects have one thing in common: models. Our interdisciplinary expertise stems from our ability to help domain experts to analyse and interpret their data, with mathematical/computations models. More specifically we use statistical models (e.g. distributions in significance analysis), stochastic models (e.g. Markov processes), differential equation models, boolean logic models, graph theoretical representations, and machine learning models.

It may sound contradictory but simple models can help explain complex systems. Modelling is an art, the art of making appropriate assumptions - leaving things out and abstract, so as to be able to explain and predict. Differential equation models are very powerful because they can **explain** mechanisms in dynamical systems. The downside is that they require sufficiently rich time course datasets. The more variables a system has, the more difficult it is to find unambiguous parameter sets. The challenge is then to get away with as many simplifying assumptions as possible, without loosing the ability to explain observations made in experiments. On the other hand, a statistical model used in significance analysis (e.g. data following a normal distribution) can only be used to determine whether there is a difference between two datasets. The datasets are easily generated for such questions. Many statistical and machine learning models are primarily focussing on **prediction**. The expertise of the 'modeller' is thus to choose the right method for a given question and dataset.

Here are some questions that are commonly asked in relations to mathematical / computational models, and which may inspire you for your article:

Machine Learning Models

- What's the trade-off between bias and variance?
- What is the difference between k-means and fuzzy c-means clustering, w.r.t. mathematical properties and pros and cons in their application?
- How is KNN different from k-means clustering?
- What is the difference between decision trees, support vector machines and regression analysis, w.r.t. to how and for what they are used?
- What is the role of the 'confusion matrix' is and how is it used to validate the accuracy of k-nearest neigbour classification.
- Which is more important to you– model accuracy, or model performance?
- Explain how a ROC curve works.
- Dimensionality reduction:
  - What are the mathematical differences between PCA, t-SNE and UMAP?
  - What are the pros and cons of principle component analysis, t-SNE and UMAP?

Dynamical Systems Models

- Why do people linearise nonlinear models?
- What models are used to explain causation?
- What does it mean for a model of nonlinear differential equations to be non-identifiable?
- Modelling a dynamical system with boolean logic and differential equations, how do these two approaches compare?
- How could the Hodkin-Huxley model, with only three differential equations, win the Nobel price for medicine?
- How can the Lotka-Volterra model be soo inaccurate and yet be so useful in science?
- By what criteria can one judge a model? (e.g. accuracy, realism, …)
- How is Occam's razor used to discuss a model's complexity in terms of simplicity and accuracy?