

Formal Languages and Automata Theory

Akash Gopinath

October 29, 2025

MMath Independent Project
Supervised by Prof. Victoria Gould



Department of Mathematics
University of York
2025 - 2026

Contents

1	What is my project	2
2	How I will achieve this	3
3	Motivations and issues I hope to address	4
4	Timeline of the project	5

Chapter 1. What is my project

My topic is on the study of formal languages and automata theory. This topic is an intersection of mathematics and theoretical computer science. It explores the relationship between algebraic structures and computational models, providing a rigorous framework for understanding the capabilities and limitations of different types of automata and the languages they recognize.

A DFA, or Deterministic Finite Automata, is a theoretical model of computation that consists of a finite set of states, an input alphabet, a transition function, a start state, and a set of accept states. A DFA can also be viewed as a generator of algebraic information. A regular language L has associated with it a monoid $M(L)$ called the syntactic monoid of L . Monoids help us determine the class of regular recognizable languages.

My first objective is to understand the algebraic structure of $M(L)$ and then use this to study varieties and pseudovarieties of monoids. A variety of algebras is a class of algebras closed under the formation of homomorphic images, subalgebras, and arbitrary direct products. When restricted to finite algebras, these are called pseudovarieties.

The next objective is to explore Eilenberg's correspondence, which establishes a bijective relationship between varieties of regular languages and pseudovarieties of finite monoids. This correspondence allows us to translate problems in language theory into problems in algebra and vice versa, providing powerful tools for analysis. Under this correspondence, a regular language L belongs to a variety of languages \mathcal{V} if and only if its syntactic monoid $M(L)$ belongs to the corresponding pseudovariety of monoids \mathcal{W} .

This connection is useful in classifying regular languages based on the algebraic properties of their syntactic monoids.

An important result in this area is Rietterman's theorem. This theorem states that a class of monoids is a pseudovariety if and only if it can be defined by a set of profinite identities. This is useful in the membership problem for pseudovarieties of finite monoids as it provides a way to characterize pseudovarieties using identities, which can be easier to work with than directly dealing with the monoids themselves.

Finally I would like to look into the membership problem for pseudovarieties of finite monoids. Which is the problem of determining whether a given finite monoid belongs to a specified pseudovariety. This problem has significant implications in automata theory and formal language theory. I would also like to look into algorithms used in the membership problem for certain pseudovarieties.

The aim of this project is to explore pseudovarieties of finite monoids and their applications in formal languages and automata theory.

Chapter 2. How I will achieve this

I have already finished reading through the relevant lecture notes provided by my supervisor. These notes cover the key concepts and techniques in formal languages and automata theory.

The next step is to delve deeper into the topic of varieties of formal languages. I am currently studying a book titled “Varieties of Formal Languages” by Jean-Éric Pin. This book provides a comprehensive overview of the theory of varieties and their applications in formal languages. This text will be the main reference for understanding how regular languages can be classified using algebraic structures.

In addition to this, I plan to consult the following textbooks/research methods:

- Samuel Eilenberg’s “Automata, Languages, and Machines” which provides foundational knowledge on automata theory and formal languages.
- John M. Howie’s “Automata and Languages” which is a gentle introduction to the subject.
- I will also look into research papers into the membership problem for pseudovarieties of finite monoids to understand the current state of research in this area.

Using literature, I will do 2 things:

- (1) **Theoretical study:** I will study the structure of the syntactic monoid of regular languages and understand how varieties and pseudovarieties are defined and classified. I will explore Eilenberg’s correspondence in detail and understand how it can be used to classify regular languages based on the properties of their syntactic monoids.
- (2) **Computational study:** I will use computational tools such as GAP (Groups, Algorithms and Programming) to experiment and compute syntactic monoids of regular languages of small DFAs. I will try to implement algorithms to test membership in certain pseudovarieties of finite monoids and analyze their efficiency and effectiveness. Through this I want to shed light on the membership problem and see if it will yield new examples/insights.

Where things might go wrong:

- Might be a bit too ambitious in terms of the scope of the project. The theoretical study might take longer than expected, leaving less time for the computational study.
- The computational study may face challenges such as the limitations of the GAP system or the complexity of the algorithms to be implemented. This could result in incomplete or inaccurate findings.

Chapter 3. Motivations and issues I hope to address

The reason I am doing this topic is because I find it interesting how algebraic structures such as monoids can be used to classify and understand formal languages. Eilenberg's correspondence is a powerful tool that connects these two areas, allowing for a deeper understanding of the properties of regular languages through the lens of algebra. For example star free languages correspond to aperiodic monoids.

This topic has both theoretical and computational aspects which I can explore which makes it appealing to me. Also after taking a course on semigroup theory last year, I developed a fascination in the subject and would like to explore its applications in automata theory.

Issues I hope to address

There are several issues I hope to address in this project:

- **Scope:** The scope of the project is quite large. It might take a lot to build up to Eilenberg's correspondence and the membership problem may be too ambitious to tackle in the time frame. Nevertheless, even with Eilenberg's correspondence alone, I can gain a lot of insight into the classification of regular languages and could do a broad survey of membership without trying to implement algorithms.
- **Computational challenges:** Implementing algorithms to test membership in pseudovarieties may be too complex in the time frame I am given. But I still hope to try and compute syntactic monoids of small DFAs using GAP since that is a finite simple process.
- **Timeframe for literature:** In the gantt chart below, I have outlined when I hope to do what task. But it is based on my current understanding of the topic. As I read more literature, I may find that some topics take longer to understand than I initially thought. It may take longer also due to lectures, exams and other commitments. I will try to manage my time effectively to ensure I can cover the necessary material within the timeframe.

Overall, I can definitely get to Eilenberg's correspondence and understand how varieties of regular languages and pseudovarieties of finite monoids are connected. And then just do a broad survey of membership of known pseudovarieties.

In testing membership, Reiterman's theorem is useful. But I will be omitting the details of the proof since the topological arguments are not in the scope of this project.

Chapter 4. Timeline of the project

Here is a Gantt chart showing the timeline for my MMath independent project:

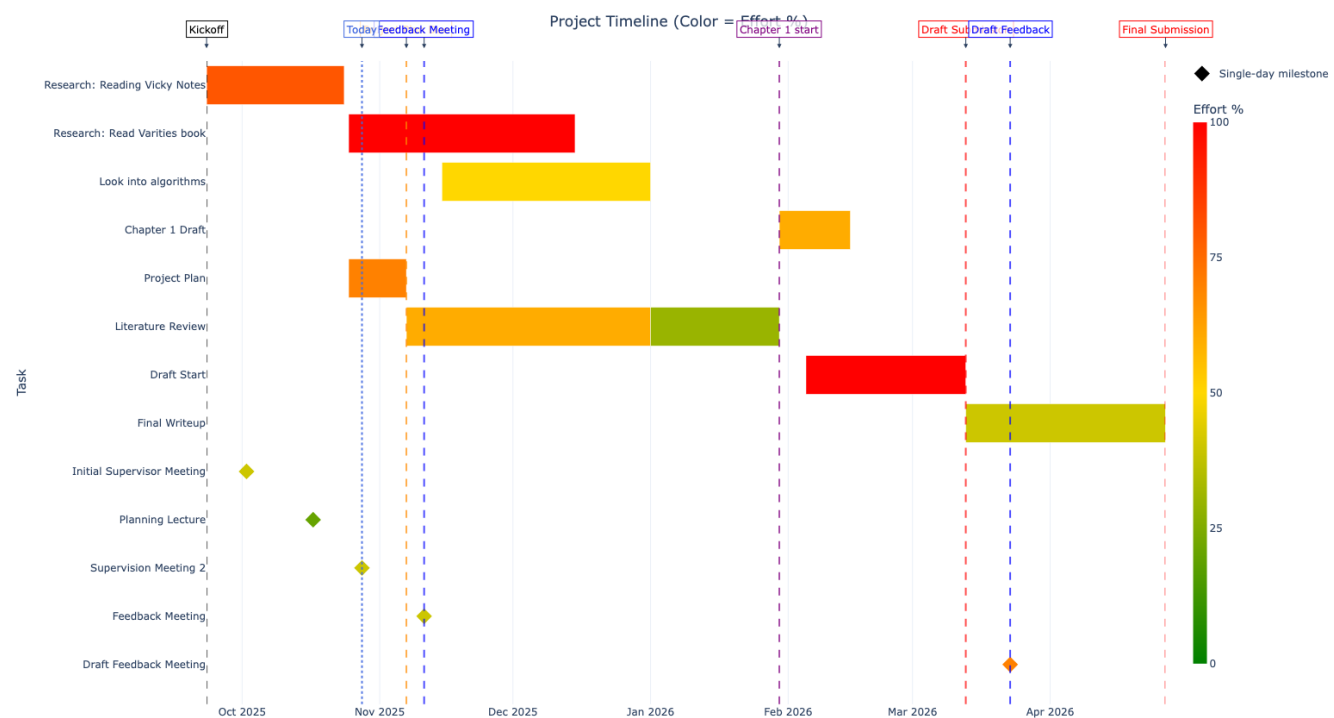
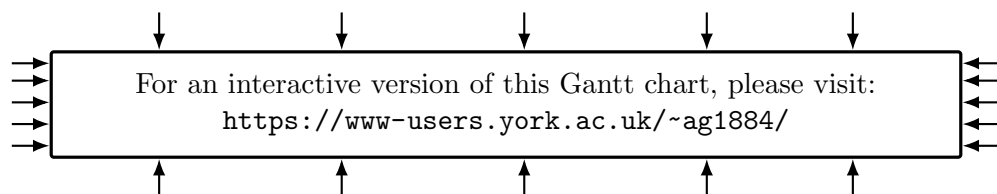


Figure 0.1: Gantt chart showing the timeline for the MMath independent project.



Written Commentary of plan

The Gantt chart in Figure 0.1 outlines the timeline for my MMath independent project. The project is structured into several key phases, each with specific tasks and milestones.

Note: The weeks mentioned correspond to the academic calendar of the University of York for the 2025-2026 academic year. Also some tasks overlap to allow for flexibility

and ensure that progress can be made even if certain tasks take longer than expected. The dates are also speculative and may change based on my progress and any unforeseen circumstances.

- **Initial Research (Weeks 1-5):** During this part, I focused on reading the lecture notes of my supervisor Prof Victoria Gould and relevant material from book "Automata and Languages" by John M. Howie. This helped me understand the basics of DFA's and syntactic monoids
- **Project Plan (Weeks 5-6):** I started working on my project plan. This also relied on me consulting additional literature in tandem.
- **Study on Varieties (Weeks 6-12):** This phase involves a deep dive into the theoretical and computational aspects of pseudovarieties of finite monoids and their applications in formal languages and automata theory. I will study Eilenberg's correspondence and related concepts. Most of this material is covered in the book "Varieties of Formal Languages" by Jean-Éric Pin and his lecture notes available online.
 - October 25-December 15: Read through relevant chapters of Jean-Éric Pin's book and understand the theoretical concepts.
 - November 15-January 1: Look into algorithms used to generate syntactic monoids of regular languages and test membership in pseudovarieties.
 - November 15-January 30th: I will also read alternative literature and look up other references.

Note: January 1-30th is reading week and exam period so I will not be doing much work on the project during this time.
- January 30-February 15: Start the draft, I will try and write up the first few chapters and theoretical background during this time.
- February 5-March 13: Start and finish draft before deadline. I will also write code to compute syntactic monoids of small DFAs using GAP. If time permits, I will also try and implement algorithms to test membership in certain pseudovarieties of finite monoids.
- March 13-April 27: Final revisions and submission. I will review the entire document, make necessary revisions, and prepare the final version for submission by the deadline. I will start the final version before the draft feedback to ensure I have enough time to make revisions based on the feedback received.