

## SOFTWARE & DATA ENGINEERING MASTER THESIS PROPOSAL

**Edoardo Riggio**

### API SCOUT: AN INFORMATION RETRIEVAL SYSTEM FOR OPENAPI SPECIFICATIONS

---

#### *Abstract*

The primary objective of this thesis is to create an information retrieval system enabling users to explore a comprehensive collection of OpenAPI specifications. The system will use historical versioned data - scraped from existing github repositories - as well as user-uploaded data.

This platform serves a dual purpose, catering to both academics and developers. In the case of academics, this platform can serve as a repository from which to conduct research, facilitating more in-depth studies on API specifications and their evolution. On the other hand, developers can look up several different examples that could help them get started with their own OpenAPI documentation.

The main contributions of this thesis include two crucial aspects: data engineering and software engineering. In the first case, the thesis will delve in the process of refining raw data and finding ways to classify the API specifications. In the second case, the thesis will define the architectural design and the development of the service itself.

---

**Contents**

**1 Introduction 2**

1.1 Context . . . . . 2

1.2 Motivation . . . . . 2

1.3 Objective . . . . . 2

**2 Proposal 2**

**3 State of the Art 2**

**4 Preliminary Results - Feasibility Study 2**

**5 Work Plan for the Spring semester 2**

**List of Figures**

**List of Tables**

# **1 Introduction**

## **1.1 Context**

## **1.2 Motivation**

## **1.3 Objective**

# **2 Proposal**

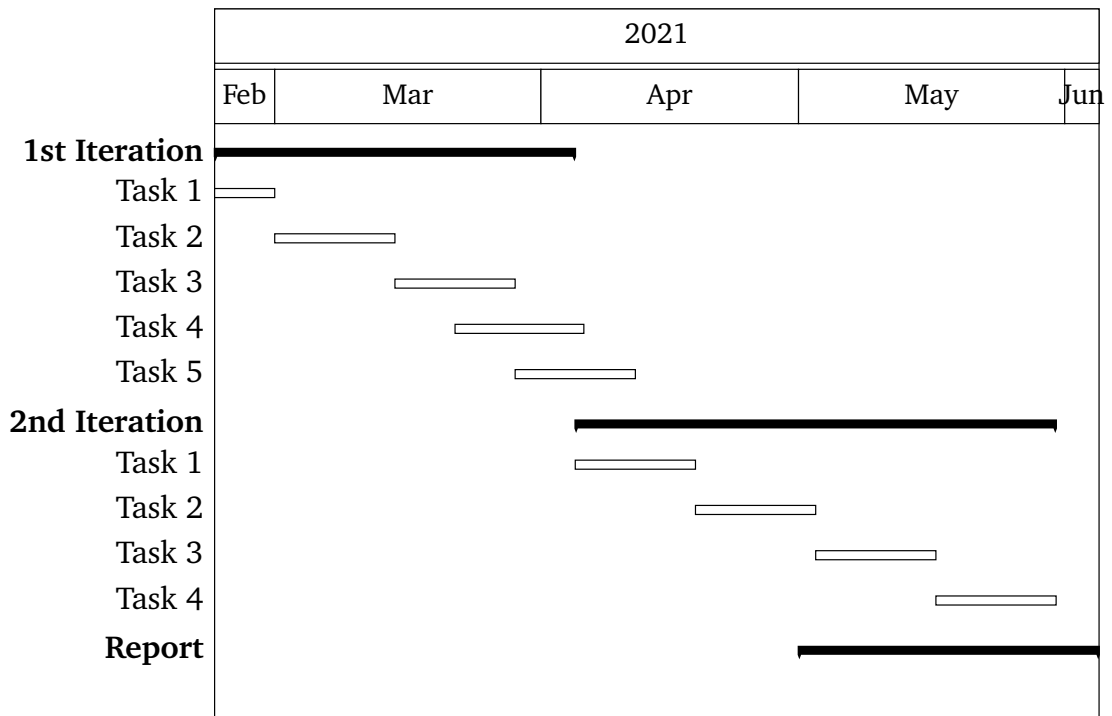
# **3 State of the Art**

# **4 Preliminary Results - Feasibility Study**

# **5 Work Plan for the Spring semester**

This section describes a high level view of the work plan for the spring semester.

- Iteration 1: Data Analysis Phase
  - Task 1: Setup project (Pick tools/technologies to use)
  - Task 2: Design of metrics to locate changes in evolution
  - Task 3: Setup database of metrics
  - Task 4: Classification of changes
  - Task 5: Evaluate and improve system
- Iteration 2: Visualization Phase
  - Task 1: High level visualization of API
  - Task 2: Drilled down visualization of API
  - Task 3: Visualizing changes
  - Task 4: Evaluate and improve system
- Iteration 3: Report



## References