



Universidade do Minho  
Escola de Engenharia

## Pedido de Admissão à Dissertação/Projeto Plano de Trabalho

<b>Ano Letivo:</b> 2021/2022	<b>Nome:</b> Hugo André Coelho Cardoso <b>Número:</b> a85006 <b>Título (em Português):</b> DataGen from Schemas <b>Título (em Inglês):</b> DataGen from Schemas
---------------------------------	--

### Enquadramento e Motivação (150-200 palavras)

O DataGen é uma ferramenta versátil que permite a prototipagem rápida de datasets e a testagem de aplicações de software. Atualmente, esta solução é das poucas que oferece tanto a complexidade como a escalabilidade necessárias para gerar datasets adequados para avaliar o desempenho de uma API de dados ou de uma aplicação complexa, permitindo testar as mesmas com um volume apropriado de dados heterogêneos.

O cerne do DataGen é a sua Domain Specific Language (DSL) que foi criada para especificar datasets, dotada de um vasto leque de mecanismos que permitem a geração de datasets bastante variados, lidando com requerimentos muito exigentes e intrincados.

O DataGen From Schemas surge como um complemento a esta aplicação, procurando gerar datasets diretamente a partir de ficheiros de especificação de vários formatos de dados (XML Schemas, JSON Schemas, etc), através das rotas aplicacionais REST expostas pelo DataGen. Desta forma, deixa de ser necessário o utilizador especificar a estrutura de dados pretendida na DSL do DataGen e basta-lhe criar um ficheiro de especificação tradicional, com os quais há um grau bastante elevado de familiarização, na generalidade. Assim, o número de use cases desta geração de datasets aumenta bastante.

### Objetivos e Resultados Esperados (150-200 palavras)

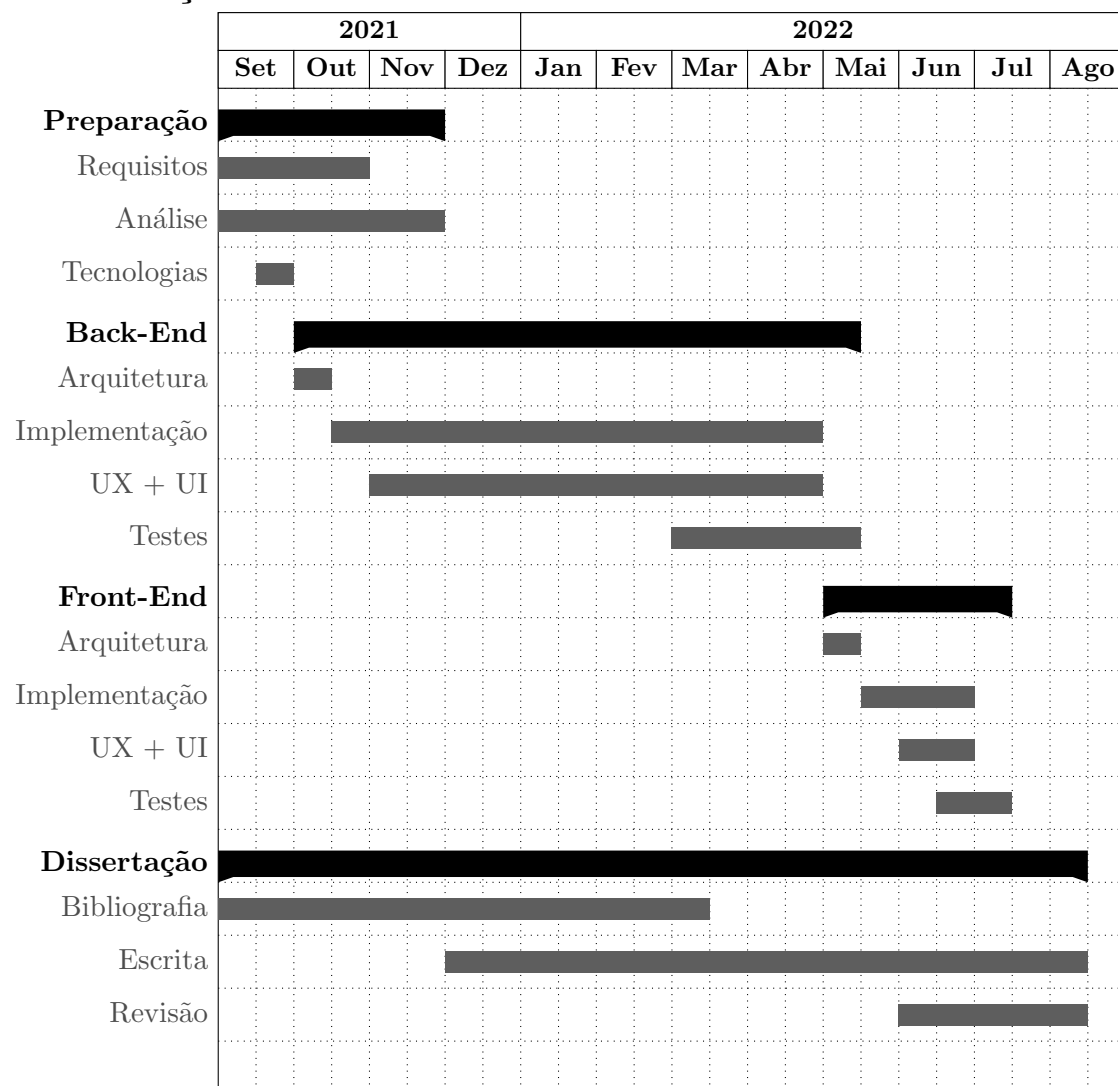
O objetivo principal da dissertação é a geração de datasets a partir de ficheiros de especificação (Schemas) de formatos de dados populares, principalmente XML.

Mais especificamente, este processo passa pelo desenvolvimento de compiladores, com base em gramáticas PEG.js, para filtrar e processar ficheiros de especificação. Cada tipo de ficheiro exige um compilador diferente, visto que os formatos de dados variam bastante em forma e não há tradução direta de uns para outros.

De seguida, é necessária a criação de um conversor capaz de gerar modelos válidos na DSL do DataGen a partir dos dados filtrados – este programa deve ser capaz de traduzir a estrutura de dados descrita no ficheiro filtrado para um modelo da DSL, que será posteriormente usado para gerar um dataset a partir das rotas aplicacionais expostas pelo DataGen.

Por fim, a criação e deployment de uma aplicação web para disponibilizar o DataGen From Schemas, com características user-friendly, documentação extensiva a explicar a sua utilização e rotas REST para possibilitar a utilização da aplicação sem recurso à interface gráfica e a eventual integração em terceiros.

## Calendarização



## Referências Bibliográficas (5-10 referências)

- [1] Antonia Bertolino et al. “Automatic Test Data Generation for XML Schema-based Partition Testing”. In: *Second International Workshop on Automation of Software Test (AST '07)*. 2007, pp. 4–4. DOI: 10.1109/AST.2007.6.
- [2] Kenneth Chiu and Wei Lu. “A compiler-based approach to schema-specific XML parsing”. In: (Jan. 2004).
- [3] Shudi (Sandy) Gao, C. M. Sperberg-McQueen, and Henry Thompson. *W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures*. English. 2012.
- [4] Zhenghong Gao et al. “A High Performance Schema-Specific XML Parser”. In: *Third IEEE International Conference on e-Science and Grid Computing (e-Science 2007)*. 2007, pp. 245–252. DOI: 10.1109/E-SCIENCE.2007.5.
- [5] Jian Bing Li and J. Miller. “Testing the semantics of W3C XML schema”. In: *29th Annual International Computer Software and Applications Conference (COMP-SAC'05)*. Vol. 1. 2005, 443–448 Vol. 2. DOI: 10.1109/COMPSAC.2005.151.
- [6] Morris Matsa et al. “A High-Performance Interpretive Approach to Schema-Directed Parsing”. In: *Proceedings of the 16th International Conference on World Wide Web. WWW '07*. Banff, Alberta, Canada: Association for Computing Machinery, 2007, pp. 1093–1114. ISBN: 9781595936547. DOI: 10.1145/1242572.1242719. URL: <https://doi.org/10.1145/1242572.1242719>.
- [7] Prajakta M. Pardeshi and Ganesh Ramachandran. “Generic method for XML generation from given XSD”. In: *2011 3rd International Conference on Electronics Computer Technology*. Vol. 5. 2011, pp. 404–408. DOI: 10.1109/ICECTECH.2011.5942030.
- [8] Filipa Alves dos Santos et al. “DataGen: JSON/XML Dataset Generator”. In: *SLATE*. 2021.

<b>Assinaturas</b>	
<b>Estudante</b>	<b>Orientador</b>
<b>Diretor do Ciclo de Estudos</b>	<b>Coorientador (se aplicável)</b>