

## Tema para Dissertação do 2º ciclo em Engenharia Informática

**Tema:** Programação paralela em Python para análise de dados

**Orientador:** Paula Prata  
(e-mail: pprata@di.ubi.pt)

Co-orientador: Maria Eugénia Ferrão (Departamento de Matemática)

### Contexto

Numa sociedade que produz cada vez maiores quantidades de dados, a análise desses dados de forma a deles retirar conhecimento útil é crucial. A análise de grandes quantidades de dados exige grande capacidade de processamento sendo a programação paralela uma via óbvia para acelerar o processo. A linguagem Python, sendo de utilização geral tem vindo a ganhar terreno em relação à linguagem R, tradicionalmente usada para análise estatística de dados.

### Objetivos

Explorar mecanismos de programação paralela da linguagem Python [1] em aplicações de análise de dados. Como caso de estudo, pretende-se analisar testes de avaliação de conhecimentos usando a teoria de resposta ao item [2] [3] e comparar com outras implementações já existentes [4].

### Tarefas

T1 – Estudar a linguagem Python e respectivos mecanismos de paralelismo.

T2 – Estudar a Teoria de Resposta ao Item.

T3 – Estudar implementações paralelas para determinar parâmetros de itens de avaliação (e.g., grau de dificuldade, grau de discriminação).

T4 – Comparar resultados com implementações sequenciais.

T5 – Escrever um artigo para publicação dos resultados.

T6 – Escrever a dissertação.

### Cronograma de Tarefas

Set 17	Out 17	Nov 17	Dez 17	Jan 18	Fev 18	Mar 18	Abr 18	Mai 18	Jun 18
T1	T1								
		T2	T2						
			T3	T3	T3				
					T4	T4			
						T5	T5		
						T6	T6	T6	T6

## Referências

- [1] <http://ipyparallel.readthedocs.io/en/latest/>
- [2] Eignor, D. R. (2007). C34B - Linking Scores Across Computer and Paper-Based Modes of Test Administration. In *Handbook of Statistics Vol 26* (Vol. 26, pp. 1099–1102). doi:10.1016/S0169-7161(06)26041-3
- [3] Lin, C. (2008). Comparisons between Classical Test Theory and Item Response Theory in Automated Assembly of Parallel Test Forms in Automated Assembly of Parallel Test Forms. *The Journal of Technology, Learning, and Assessment*, 6(8), 1–43. Retrieved from <http://www.jtla.org>
- [4] Rizopoulos, Dimitris. "Irm: An R Package for Latent Variable Modeling and Item Response Analysis." , *Journal of Statistical Software* [Online], 17.5 (2007): 1 - 25. Web. 9 Jul. 2017.