# *ExtremeAlert – Deteção de Extremismo e Violência em Texto Social*

Proposta de Projeto

Orientador: João Paulo Cordeiro

Ano Letivo de 2018/2019

## 1 Objetivos

As novas tecnologias também vieram abrir novas possibilidades a pessoas e grupos mal-intencionados e com perfile potencialmente violento. Por exemplo, é sabido que os grupos jihadistas usam fóruns na Dark Web para doutrinar e recrutar novos elementos, com apelo frequente ao uso de violência [1, 2]. Cada vez mais, as forças de segurança e as unidades de contra terrorismo dependem de sistemas automáticos sofisticados, capazes de detetar conteúdo textual ameaçador à segurança dos cidadãos. Neste contexto, técnicas avançadas de *Processamento de Linguagem Natural* (PLN) e *Prospeção de Texto* (Text Mining) coordenadas por *Inteligência Artificial*, tornam-se indispensáveis para uma deteção, monitorização e prevenção eficaz da perigosidade, muitas vezes latente e invisível a olho nu.

Este trabalho pretende fazer uma contribuição na direção da criação deste género de sistemas, almejando a construção semiautomática de um "léxico crítico" (LC), i.e. um léxico preciso e completo dos termos e expressões mais comprometedoras e que ajudem a identificar texto que contenha traços de violência e extremismo. Numa primeira fase, usaremos uma coleção extensa de textos deste género de onde se extrairão os termos iniciais do LC, usando *modelação de tópicos* com LDA [3, 4]. Numa segunda fase, expandiremos a coleção inicial de termos LC usando técnicas de *Word Embeddings* (e.g. *word2vec*) [5]. Numa terceira e última fase, testaremos o LC, na implementação/treino de classificadores capazes de identificar conteúdo crítico, em novo texto social.

## 2 Plano de Trabalho

O desenvolvimento deste projeto deve seguir a seguinte ordem:

**T1**: Preparação da coleção de textos a usar;

**T2**: Estudo das técnicas e bibliotecas de *LDA* e *Embeddings*;

**T3**: Construção iterativa e automática do Léxico Crítico.

**T4**: Exploração de classificadores para identificação de texto crítico;

**T5**: Escrita do relatório de projeto.

## 3 Cronograma

**T1**: duas semanas (13%).

**T2**: três semanas (21%).

**T3**: três semanas (21%).

**T4**: quatro semanas (24%).

**T5**: três semanas (21%).

## 4 Requisitos Técnicos / Académicos

O aluno deve possuir boas competências em domínios fundamentais, tais como Programação e Inteligência Artificial, devendo também estar preparado e disposto a explorar novas tecnologias.

## 5 Resultados Esperados

- O Léxico Crítico, identificador de texto violento;
- Classificadores para identificação de texto violento;
- O relatório do projeto.

## 6 Referências


[1] Scanlon, J. R., & Gerber, M. S. (2014). Automatic detection of cyber-recruitment by violent extremists. Security Informatics, 3(1), 5.

[2] Scanlon, J. R., & Gerber, M. S. (2015). Forecasting violent extremist cyber recruitment. IEEE Transactions on Information Forensics and Security, 10(11), 2461-2470.

[3] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

[4] Graham, S., Weingart, S., & Milligan, I. (2012). Getting started with topic modeling and MALLET. The Editorial Board of the Programming Historian.

[5] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119)..


## 7 Contactos


*João Paulo da Costa Cordeiro* (`jpaulo@di.ubi.pt`)
UBI — Departamento de Informática, Gabinete 4.3.