

# *SentiTerms – Categorias de Sentimentos e Termos Salientes*

Proposta de Projeto

Orientador: João Paulo Cordeiro

Ano Letivo de 2018/2019

## 1 Objetivos

Este trabalho enquadra-se na área de Análise Automática de Sentimentos em Texto (AAST) [1,2], uma subárea do *Processamento da Linguagem Humana* que tem vindo a ganhar uma elevada importância, com múltiplas aplicações práticas. Sentimentos e emoções são fenómenos humanos complexos, exigindo uma análise atenta e detalhada de modo a evitar conclusões incompletas, desadequadas e mesmo falaciosas. Mais recentemente, têm sido considerados modelos mais ricos de representação de emoções, como é o caso do *Big Five* [3], do *Circumplex* [4] e do modelo de Plutchick, das oito emoções [5].

Em termos de recursos linguísticos, as abordagens computacionais para AAST baseiam-se essencialmente no uso de léxicos de sentimento (e.g. SentiWordNet [6]), nos quais cada termo tem associado um valor de sentimento. Todavia, estes léxicos são ainda recursos pouco sensíveis e mesmo desligados dos modelos mais ricos, referidos, tendendo essencialmente a determinar a polaridade de um termo: negativo, neutro ou positivo. Há uma necessidade de proceder a um estudo do modo como os léxicos de sentimento se distribuem pelas várias dimensões do campo emocional humano.

Assim, este trabalho preconiza um estudo empírico e computacional do modo como os léxicos se distribuem pelas dimensões emocionais humanas. No entanto, em vez de partir de modelos teóricos pré-concebidos, vamos partir dos dados (léxicos) e explorar agregação (clustering) de termos, com base em texto de redes sociais, para posteriormente descobrir os modelos, ou pelo menos ver até que ponto os que existem se encaixam com a realidade dos dados.

## 2 Plano de Trabalho

O desenvolvimento deste projeto deve seguir a seguinte ordem de trabalho:

**T1:** Estudo da área de AAST;

**T2:** Exploração de léxicos existentes;

**T3:** Definição de características estratégias para a aglomeração;

**T4:** Ciclo experimental: implementação, medição e melhoria;

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

### 3 Cronograma

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

**T2:** duas semanas (13%).

**T3:** duas semanas (13%).

**T4:** seis semanas (40%).

**T5:** três semanas (21%).

### 4 Requisitos Técnicos / Acadêmicos

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

- Descoberta e caracterização de aglomerados de léxicos de sentimento;
- Um novo recurso linguístico para AAST;
- O relatório do projeto.

### 6 Referências

- [1] Taboada, M. (2016). Sentiment analysis: an overview from linguistics.
- [2] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- [3] John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999), 102-138.
- [4] Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6), 1161.
- [5] Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.) *Emotion: Theory, research, and experience* (Vol. 1, pp. 189-217).

- [6] Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec* (Vol. 10, No. 2010, pp. 2200-2204).

## 7 Contactos

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