



sherpa.ai

Building Automatic Trending News Storylines

AI Research & Development

Master Thesis Project

OCTOBER 2019

Project Description

There are two aspects of news recommendation systems that are very important for a successful user experience, the detection of trending news and the identification of groups of news with the same or similar content. **The aim of this project is to make use of unsupervised classification techniques to reconstruct a trending news storyline over time.**

Given a trending new, the goal is to explore efficient unsupervised classification techniques to exploit pairwise text similarity distances to detect similar and related past news, and later on, reconstruct the trend storyline. The analysis could be performed over a finite time, i.e., last week, or be opened to the whole ancillary information.

Working Plan & Expected Results

1. Exploratory analysis of the trend news Sherpa's benchmark dataset and the similarity distances
2. Experimental design of finite time windows short-term experiments
3. Experimental design of opened time windows long-term experiments
4. Report

Academic and Industrial Mentoring

- Dr. Miguel A. Veganzones (Sherpa AI Director)
- Prof. Jose A. Lozano (Intelligent Systems Group, EHU/UPV)

Candidate Profile

Basic knowledge of:

- Statistics
- R / Python

Interest on:

- Machine Learning
- Natural Language Processing

Benefits and Practical Information

- Funding: 650€ / Month
- Duration: 3 – 6 Months
- Location: Aula SHERPA, Fac. Informática San Sebastián

Sherpa Europe, S.L. (Sherpa) accepts no liability for the content of this document, or for the consequences of any actions taken on the basis of the information provided. This document is intended to provide preliminary guidance in anticipation of further discussion and has not been prepared with the level of due diligence and analysis that would be needed to constitute a commitment of Sherpa. Anyone who receives this document are cautioned to consider that its contents are unaudited and that it may contain inaccurate, incomplete or summarized information, which may be change without notice. This document is confidential and its contents may not be totally or partially disclosed or reproduced, without the prior written consent of Sherpa. By allowing you access to the aforementioned document we shall have no liability, duty or obligation of any kind to you.