



PhD position in text stream analytics
Carnegie Mellon University | Universidade NOVA Lisboa
Dual-Degree Doctoral Program

<https://www.lti.cs.cmu.edu/learn>

Application deadline: 15 December 2016

We invite applications for a fully funded 5-year PhD position with the GoLocal project. The project develops media monitoring technology and context aware recommendation systems. The velocity and variety of today's data streams demand new live indexing and search techniques. The applicant will investigate stream analytics and filtering to monitor multiple real-time social-media streams.

The position offers the opportunity to do research in a diverse, international team. The successful applicant will spend two years at Carnegie Mellon supervised by Prof. Jamie Callan and three years at Universidade NOVA Lisboa supervised by Prof. Joao Magalhaes; and will receive a PhD from both Carnegie Mellon University's School of Computer Science and the Faculdade de Ciências e Tecnologia of the Universidade NOVA Lisboa.

The applicant must have an BSc in Computer Science or an equivalent degree; have a good background in text information processing and machine learning; have strong programming experience in languages such as Python, Java, and C++; and be fluent in English.

Applicants must submit two recommendation letters and TOEFL and GRE scores. For details about the application process, please see <http://www.cmuportugal.org/tiercontent.aspx?id=114>.

Please visit <https://www.lti.cs.cmu.edu/learn> for more details about the Dual-Degree program.

This position is at Carnegie Mellon University / Universidade NOVA Lisboa, and is supervised by Prof. Jamie Callan (callan@cs.cmu.edu) and Prof. João Magalhães (jmag@fct.unl.pt).

GoLocal Project: <http://novasearch.org/projects/golocal/>

Nowadays, streams of Web user data are mostly discarded by current Web information systems. User location, devices, services and other sensors hide specific information consumption patterns that could be identified by online services to better answer consumer needs. Most of this data is only useful during a short period of time and is related to short-lived events, far shorter than the time a batch and non-distributed data mining algorithm needs to timely process large-scale data. In this context, the project will target two technological goals:

1. We will investigate **media monitoring technology** to track the popularity or reputation of entities on the Web. Knowing the right market value of a brand or a product is a valuable information with many uses.
2. The second technological objective concerns **context-aware recommendation**. We propose to innovate in this area by investigating new ways of inferring clues from the user context and by compiling a set of items to recommend to groups of users.