

# Introducción al Procesamiento del Lenguaje Natural

Dr. Eduardo Xamena  
{examena@unsa.edu.ar}

Mayo 2017

## 1. Resumen

El Procesamiento del Lenguaje Natural (PLN) es un campo de la ciencia que combina diferentes técnicas computacionales derivadas de la Inteligencia artificial, el Aprendizaje automático y la Inferencia estadística (Murphy, 2012) con la lingüística aplicada (Manning et al., 1999), con el propósito de lograr la comprensión y el procesamiento de textos por parte una computadora. Hoy en día existen diversas herramientas de software que nos permiten llevar a cabo esta tarea (Cunningham et al., 2011; Bird et al., 2009). Un ejemplo de estas herramientas es GATE (Cunningham et al., 2011), la cual provee un set muy completo de algoritmos para la implementación del PLN. Entre las distintas aplicaciones que podemos encontrar para el PLN podemos mencionar: clasificadores automáticos de textos en categorías, programas traductores de documentos de un lenguaje a otro, paquetes de reconocimiento de entidades nombradas en textos, etcétera.

El propósito de este taller es mostrar la implementación de la herramienta GATE para PLN, utilizando un corpus de documentos relacionados a la investigación científica. Se mostrará el desarrollo de una aplicación JAVA que emplee las funciones de GATE para obtener distintas informaciones de los documentos del corpus.

## 2. Objetivos

- Que el alumno comprenda los conceptos del PLN y sus aplicaciones
- Introducir al alumno en el manejo de la librería GATE para PLN en un caso concreto
- Que el alumno sepa interpretar los resultados arrojados por la plataforma respecto al análisis de textos

## 3. Contenido

- Introducción a PLN
- La Herramienta GATE
  - Funcionalidad básica de la herramienta
  - Aplicación a un corpus concreto
  - Análisis de los resultados

## 4. Consideraciones sobre el taller

- Destinatarios: alumnos avanzados, egresados y docentes de la carrera Informática.

- Las prácticas se realizarán en PC/Notebook y se utilizará Eclipse/Java y la librería GATE<sup>1</sup>.

## Referencias

Bird, S., Klein, E., y Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. "O'Reilly Media, Inc."

Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N., Roberts, I., Gorrell, G., Funk, A., Roberts, A., Damljanovic, D., Heitz, T., Greenwood, M. A., Saggion, H., Petrak, J., Li, Y., y Peters, W. (2011). *Text Processing with GATE (Version 6)*.

Manning, C. D., Schütze, H., et al. (1999). *Foundations of statistical natural language processing*, volume 999. MIT Press.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

---

<sup>1</sup><https://gate.ac.uk/>