

# Aplicaciones de Procesamiento de Lenguaje Natural y Ciencia de Datos



**Anabella De Battista, Soledad Retamar, Patricia Cristaldo, Esteban Schab, Carlos Casanova, Lautaro Ramos, Ramiro Rivera, Cristhian Richard, Lucas La Pietra, Lautaro Retamar, Franco Miret, Rodrigo Mignola, Juan Manuel Franzante**  
 Grupo de Investigación en Bases de Datos, Departamento de Sistemas de Información. UTN – FRCU. Entre Ríos, Argentina  
 {debattistaa, retamars, schabe, ramosl, riverar, richardc, cristaldop, lapietral, retamarl, miretf, mignolar, franzantejm}@frcu.utn.edu.ar

**Norma Edith Herrera**  
 Departamento de Informática  
 Univ. Nac. de San Luis, Argentina  
 {nherrera}@unsl.edu.ar

## Contexto

Las aplicaciones que se presentan en este trabajo se desarrollan en el ámbito del proyecto de investigación Descubrimiento de conocimiento en Bases de Datos (C. 5109) del Grupo de Investigación en Bases de Datos, perteneciente al Departamento Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay.

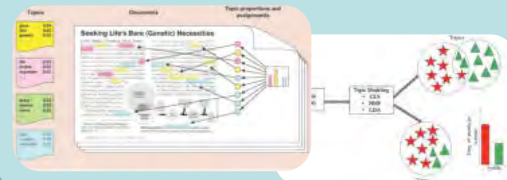
## Líneas de Investigación

La línea de trabajo principal del proyecto de investigación es el estudio, análisis y comparación de técnicas de minería de datos para el tratamiento de textos en español, el análisis de su desempeño en distintos escenarios y la propuesta de modificaciones o mejoras a las técnicas de minería de textos existentes para incrementar la calidad de los resultados en el tratamiento de textos en español. También está previsto realizar una evaluación del funcionamiento de técnicas de minería de datos sobre conjuntos "tradicionales" enriquecidos con atributos provenientes de textos relacionados. Durante el año 2020 se desarrollaron aplicaciones de Ciencia de Datos orientadas a: seguimiento de la pandemia de COVID-19 y de la epidemia de Dengue, aplicación de Topic Modelin a un corpus de noticias en español, aplicación de Análisis de Sentimientos a un corpus de comentarios de usuarios recolectados de la plataforma TripAdvisor. Actualmente se trabaja en analítica de datos aplicada al deporte, en particular al básquet.

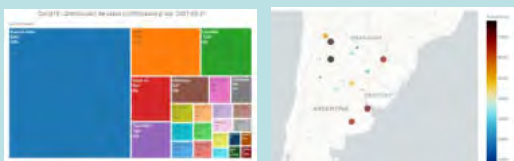
### Análisis de Sentimientos



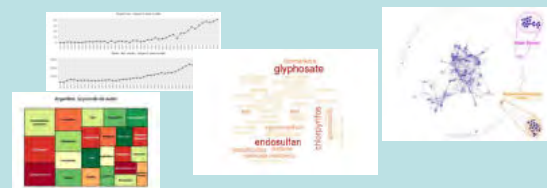
### Topic Modelling



### Ciencia de Datos



### Visualización de datos



## Resultados obtenidos y esperados

Con este proyecto se espera lograr propuestas novedosas de técnicas de minería de textos aplicables a textos en español y el desarrollo de soluciones en el ámbito de la Ciencia de Datos. Como marco metodológico para el desarrollo de los proyectos se aplica una metodología ágil para proyectos de ciencias de datos propuesta en el marco del grupo de investigación. Se han firmado convenios con instituciones públicas y empresas privadas para el desarrollo de actividades de transferencia.

## Formación de Recursos Humanos

En el marco del proyecto se está desarrollando una tesis de maestría y una de las investigadoras está cursando el Doctorado en Informática. Se cuenta con un becario graduado con beca de iniciación a la investigación y dos becarios alumnos de la carrera Ingeniería en Sistemas de Información que inician su formación en la investigación. Está prevista la realización de al menos una Práctica Supervisada de Ingeniería en Sistemas de Información en el marco del proyecto.